Author:
Steventon, Joe

Title:
Development and optimisation of a protocol for the extraction of microbial DNA from clinical pulmonary samples

University of BRISTOL

Development and optimisation of a protocol for the extraction of microbial DNA from clinical pulmonary samples

Name: Joseph Steventon

School: Cellular and Molecular Medicine

Supervisor: Professor Ruth Massey

Submission date: April 2020

A thesis submitted to the University of Bristol in accordance with the requirements for award of the degree of Master of Science by Research (MSc(R)) in the Faculty of Life Sciences

i. Abstract

Ventilator-associated pneumonia (VAP) is the most common nosocomial infection in patients admitted to the ICU and a leading cause of mortality in critically ill patients. The rise in antimicrobial resistance threatens current treatment methods, and if not addressed will lead to a significant increase in mortality. Rapid and accurate diagnostic methodologies are needed to combat this issue. To develop and test the potential of these DNA based technologies, we focus here on *Staphylococcus aureus* as it is a common cause of antimicrobial resistant VAP.

The first step in the application of any DNA-based test is the reliable extraction of the bacterial DNA. In this project three DNA extraction methods were compared for accuracy and sensitivity: a commercially available DNA extraction kit, and two methods which use traditional phenol/chloroform methods for DNA purification but differ in their lysis technique; one using physical bead-beating and the other using MetaPolyzyme, a multi-lytic enzyme mix.

To test each method's ability to extract *S. aureus*, extracted DNA was amplified using qPCR with generic 16S rRNA gene primers and *S. aureus* specific *nuc* gene primers to determine the relative quantity of *S. aureus* within a sample. On pure broth grown cultures the bead-beating method was determined as the most effective and was subsequently developed and optimised for use on pleural fluids provided to us by the NHS. With these more complex materials the use of phase lock gel (PLG) reduced the variability observed when using phenol/chloroform in standard laboratory tubes.

Having established the optimal extraction methodology, and alongside control plasmids, the limit of detection was determined for two sets of PCR primers, a generic bacterial 16S set and a *S. aureus* specific *nuc* set. The limit of detection of spiked pleural fluids was ~$1x10^4$ CFU/ml using the 16S primer set and ~$1x10^3$ CFU/ml for the *nuc* set. As this work is developed further its use in a clinical diagnostic setting will be evaluated to provide rapid diagnoses of pleural infections.

ii. Acknowledgements

Firstly, I would like to thank my supervisor, Professor Ruth Massey, for her patience and guidance in overcoming the various challenges that I faced during the completion of this project, both scientific and personal.

I would like to acknowledge Dr Michael Cox for providing the original protocol for one of the DNA extraction methods used in this project, the p(16S) plasmid standard, and the oligonucleotide sequences for the 16S rRNA primers; John Shaw for providing the pCDF-Duet-1 plasmid vector used to produce the p(*nuc*) plasmid standard; and Dr Fergus Hamilton for providing the clinical samples.

Thank you to the D57 and D60 lab groups for their support, advice and friendship throughout this project. Especially to the Massey group; Emily, Seána, Tarcisio, Daniel, Ed, Dora, Alaa and Dina. Thank you for always being happy to share and build on ideas, always helping to troubleshoot any experimental complications, and always providing an enjoyable working environment.

Finally, I would like to thank my family and Nikita for their unfailing support. Thank you for pretending to listen when I talk about science, for the continuous encouragement, and for offering to read my thesis until you found out how long it was. I appreciate it all.

iii. Author’s declaration

I declare that the work in this dissertation was carried out in accordance with the requirements of the University's Regulations and Code of Practice for Research Degree Programmes and that it has not been submitted for any other academic award. Except where indicated by specific reference in the text, the work is the candidate's own work. Work done in collaboration with, or with the assistance of, others, is indicated as such. Any views expressed in the dissertation are those of the author.

Signed: Joseph Steventon **Date:** 09/04/2020

Word count: 26,228

Table of Contents

iv. Table of figures

v. Table of tables

vi. Abbreviations

A260/280	Ratio of absorbance at 260nm vs 280nm
AAS	Aluminium ammonium sulphate
ATP	Adenosine triphosphate
BAL	Bronchoalveolar lavage
BB+P/C	Bead-beating + phenol/chloroform
CFU	Colony forming units
CI	Chloroform:isoamyl alcohol
Cq	Quantification cycle (also known as Ct: cycle threshold)
CTAB	Hexadecyltrimethylammonium bromide
dNTP	Deoxyribonucleotide triphosphate
dPCR	Digital PCR
EDTA	Ethylenediaminetetraacetic acid
ENB	Electromagnetic navigational bronchoscopy
gDNA	Genomic DNA
IUPAC	International Union of Pure and Applied Chemistry
LB	Luria-Bertani
LME	Lysing Matrix-E
LoD	Limit of detection
LPA	Linear polyacrylamide
Met+Kit	MetaPolyzyme + DNA extraction kit
Met+P/C	MetaPolyzyme + phenol/chloroform
mic	Magnetic induction cycler
NTC	Non-template control
OD	Optical density
PBS	Phosphate-buffered saline
PCI	Phenol:chloroform:isoamyl alcohol
PCR	Polymerase chain reaction
PEG	Polyethylene glycol
PF	Pleural fluid
PLG	Phase lock gel
QACs	Quaternary ammonium compounds
qPCR	quantitative PCR
RPM	Revolutions per minute
rRNA	Ribosomal RNA
SD	Standard deviation
SOP	Standard operating procedure
T_a	Annealing temperature
TE	Tris-EDTA
T_m	Melting temperature
TNTC	Too numerous to count
Tris	Tris(hydroxymethyl)aminomethane
TSA	Tryptic soy agar
TSB	Tryptic soy broth
URI	University of Rhode Island

1. Introduction

1.1. Pneumonia

Pneumonia is defined as inflammation of the tissue in the lungs, usually due to a bacterial infection [1], the swelling causes areas of the lung to become consolidated and fill with liquid, causing loss of function. The healthy lung contains complex communities of commensal microorganisms with microbe-microbe and microbe-host interactions contributing to the maintenance of the healthy state of the lung; commensals are thought to confer protection from respiratory pathogens by inducing a constant, low-scale response from the host's immune system, preventing an infection from being established [2]. However, competition between commensal organisms can lead to the development of infection by one of these commensal species, blurring the boundary between commensal and pathogen [3]. The complexity of the human lung microbiome is known to be intricately involved in the development of pneumonia [4], however there is currently a significant gap in this area of research [3].

1.2. The lung microbiome

Microbiome research is one of the fastest growing areas of cutting-edge investigation and our understanding of the complex relationships between human and microbial life is rapidly expanding. Recent advances in molecular techniques, such as the advent of next generation sequencing, have enabled a new depth of understanding in this area. For many years, the healthy human lung was considered a sterile organ [5]; this inference was likely due to the difficulties in examining the composition of the lung exclusively through culture-dependent techniques. The complexity of the lung environment provides an example of an in-host location creating a niche for many viable but not culturable organisms [6, 7]. Replicating a culturable environment for these organisms is difficult due to the range of conditions found within the lung, from the oxygen availability gradient, to pH, temperature and nutrient availability gradients, all of which provide selective pressures influencing the precise microbial communities able to occupy this niche. Furthermore, host immune factors change with distance into the lower respiratory tract, as does local competition of microbial life found here [8]. The culmination of all these factors leads to a difficult environment to simulate *in*

vitro. Advances in culture independent techniques enable us to remove a sample from any location within the lower respiratory tract, isolate the total DNA within that sample and analyse its composition through qPCR and rapid next-generation sequencing techniques. Genetic-based approaches have shown that the human lung contains approximately 2.2 x 10^3 bacterial genomes per cm^2 (assuming 1 copy of the 16S rRNA gene per genome) [9], with most microorganisms present belonging to four phyla: Bacteroidetes, Firmicutes, Proteobacteria, and Actinobacteria [10]. For a typical, healthy individual, the most common genera present in the lungs are *Corynebacterium, Prevotella, Staphylococcus, Streptococcus, Veillonella, Haemophilus* and *Neisseria* [9]. The discovery that the lungs do have a distinct, natural microbiome has prompted further research into the dynamics of the microorganisms residing in this highly dynamic environment.

Diseased states are strongly correlated with dysbiosis, or imbalance, of the natural pulmonary equilibrium [11-13]. One of the leading models for the ecology of the lung microbiome is the adapted island model [14] (Figure 1); this model is based on a well-established model of island ecology and suggests that the microbial movement within the lungs follows similar patterns to migration of animals between land masses. The model states that a main population resides on the mainland, and over time, individuals will migrate from the mainland to the islands. The closer an island is to the mainland, the higher the likelihood of migration, resulting in a stratified series of populations on islands where each resembles the mainland more similarly than subsequent, more distant islands. This directly impacts species richness in each location, with species richness decreasing with distance from the main population, and every species present at the furthest island must have passed by the closer islands first.

The adapted island model adjusts this theory to fit the respiratory tract; the upper tract takes the place of the mainland and is the source of all species found in the lower tract, which takes the place of the islands. For any given site along a healthy respiratory tract, microbial species richness is mostly determined by immigration and elimination of microbes originating in the upper respiratory tract. The lung microbiome is maintained through an equilibrium of immigration via microaspiration of microorganisms in the air, and elimination through standard mucociliary action

and coughing. This allows a steady change in the microbiome over time, whilst maintaining a core lung microbiome of common microorganisms. If this equilibrium is compromised, regional growth conditions become a more significant factor for determining the composition of the microbiome. Dysbiosis of the lungs can produce an opening for opportunistic pathogens to establish an infection, leading to inflammation of the lungs and the development of pneumonia.

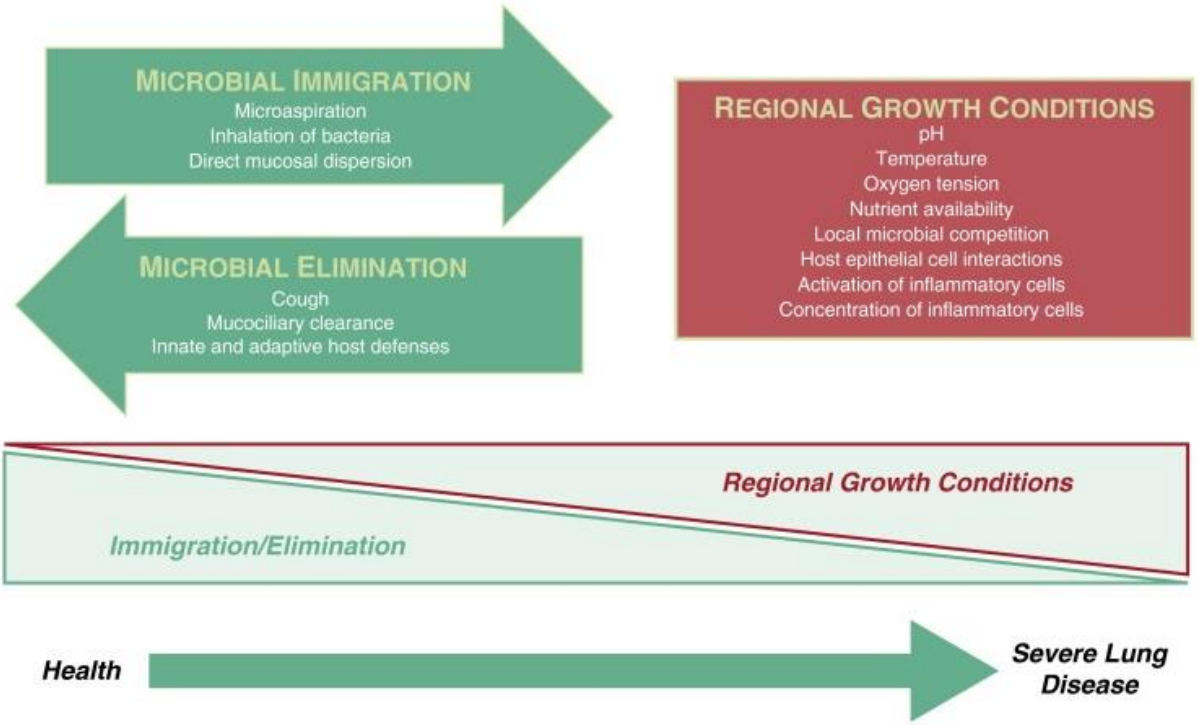


Figure 1 – *Adapted island model of the respiratory microbiome: The three factors which determine the respiratory microbiome (immigration, elimination, and relative growth of organisms) and their relationship with health and disease. (Source: Dickson et al., 2015)[14].*

1.3. Ventilator-associated pneumonia

Any condition resulting in a patient becoming unable to breathe unassisted requires artificial ventilation, this is one of the largest changes which can occur in the lungs, significantly impacting the environment for the commensal organisms present. If pneumonia develops more than 48 hours after intubation, it is termed ventilator-associated pneumonia (VAP) [15]. The effect of intubation on the lung microbiome is not fully understood, however intubation is highly likely to affect the equilibrium between microbial immigration and elimination, and the ventilator itself provides a new

environment which can be colonised. Any changes to the composition of commensal microorganisms present in the lung may enable the establishment of an opportunistic infection; it has been shown that the level of dysbiosis in the lungs following intubation is more profound in patients who develop VAP compared to those who do not develop pneumonia [16].

Microbial load of the lung is typically determined from samples of bronchoalveolar lavage (BAL) fluid, and pneumonia is diagnosed when the load rises above a threshold of 10^4 colony forming units per ml (CFU/ml) [17]. BAL is collected by introducing 130 to 150ml of sterile saline into the lungs before collecting and examining the contents through traditional culture methods. It is considered the gold standard for research into microbial communities in the lower respiratory tract as BAL fluid has been shown to have the same microbial composition as the lung tissue microbiota itself [18].

Pneumonia is typically treated using antibiotics to clear the infection, with VAP accounting for almost half of all antibiotics given in intensive care units (ICU's) [19, 20] due to it being the most common nosocomial infection in patients admitted to the ICU [16] and a leading cause of mortality in critically ill patients [21]. Each year, between 10,000 and 20,000 patients in the UK [22] and between 250,000 and 300,000 patients in the USA are diagnosed with VAP [23]. Clinical reliance on antibiotics for treatment combined with the increasing prevalence of antimicrobial resistance has significantly impaired our ability to treat these infections effectively. In order to adapt treatment strategies to be more efficient and sustainable, further research must be carried out to determine the dynamics of the opportunistic pathogens responsible for VAP in the periods prior to and during diseased states.

1.4. Staphylococcal pneumonia

Staphylococcus aureus is consistently one of the most important causes of nosocomial infection and VAP [24, 25]. It is one of the ESKAPE pathogens (*Enterococcus faecium*, *Staphylococcus aureus*, *Klebsiella pneumoniae*, *Acinetobacter baumannii*, *Pseudomonas aeruginosa* and *Enterobacter* species) which have been identified as the leading causes of nosocomial infection throughout the world [26]. This opportunistic pathogen is highly successful in establishing infection

due to a number of important virulence factors, including staphylococcal protein A (SpA) which is one of the most abundant proteins on the surface of *S. aureus* [27] and mediates binding to a membrane protein on the lung epithelium, Tumour Necrosis Factor Receptor 1 (TNFR1) [28]. This binding leads to a severe inflammatory response in the lungs which is a hallmark of pneumonia. Furthermore, the increase in antibiotic resistance is a severe complication for the treatment of bacterial infections. After a patient has received antibiotic treatment, if they then develop VAP it is far more likely that the causative strain will be resistant to further antibiotics [29], significantly increasing the likelihood of mortality.

As a major threat to public health and one of the most commonly reported pathogens associated with ventilator-associated pneumonia (VAP) [30], *S. aureus* pneumonia is an ideal model to use when investigating the dynamics of an opportunistic pathogen throughout the entire infection-to-disease process. One potential caveat of using staphylococcal species in studies such as this is that it is notoriously difficult to accurately and consistently extract DNA from [31, 32]. Difficulties in staphylococcal DNA extractions are due this organism's complex cell wall and its ability to form biofilms which can prevent chemical and biological agents for cellular lysis from reaching all cells in a population. Therefore, if the DNA extraction technique used is sensitive enough to accurately and consistently extract DNA from a difficult to lyse organism such as *S. aureus*, then the DNA from all other species within the sample should also be extracted, giving a representative view of the sample composition.

To further research on the dynamics of staphylococcal pneumonia, it is essential to have an accurate and reliable method for extracting DNA from pulmonary clinical samples, such as BAL. This will be vital as culture independent techniques become the new paradigm for microbial research, especially in the context of the microbiome. Microbiome studies commonly use the 16S rRNA gene as a genetic barcode to identify species. To understand the dynamics of MRSA within a microbiome environment, it will be essential to extract all microbial DNA and assess the composition, with specific regard to the proportion of DNA within each sample which is staphylococcal in origin. In a longitudinal study, this would allow observation of the dynamics of this organism in the reference frame of the wider lung microbiome.

It is especially important to be able to collect an accurate representation of organisms with low prevalence in the lung microbiota at the beginning of longitudinal studies in order to give a clear picture of how the dynamics have changed. Determining whether the organism which causes the potentially lethal pneumonia was present in the patient's lung upon hospital entry or if it was acquired whilst in care could be a vital piece of information when determining the correct course of clinical care to prevent future infections. This also has implications on the stewardship of antibiotics which may treat non-life-threatening infections whilst simultaneously enabling antimicrobial resistant opportunistic pathogens to establish subsequent infections.

1.5. DNA extraction methods

It is important that DNA extraction from clinical samples of unknown composition must be as comprehensive as possible, enabling extraction of DNA from all microbes within a sample with as high efficacy as possible. Imbalances in extraction efficiencies at an early stage of any project reliant on genetic techniques will lead to erroneous results. Once the total DNA has been extracted from clinical pulmonary samples, the composition of the sample can be determined by amplifying the 16S rRNA gene which is universal to all bacterial species. Small variations within the variable regions of the 16S rRNA gene allow it to be used as a barcode to determine which species a specific sequence originates from. The 16S rRNA gene can be amplified using traditional PCR methods, and the amplified product sequenced to determine the species present within a sample. qPCR is a type of PCR which enables quantification of a product throughout the amplification cycle by detecting the level of a fluorescent signal produced by the product amplified during a reaction; this enables quantification of the organisms present in a sample based on their 16S copy number when compared to a standard of a known quantity.

There are many different methods by which DNA can be extracted, however methods can be biased for or against groups of organisms and can introduce contaminants which may affect downstream applications [33]. DNA extraction methods include applying physical force, enzymes, detergents, heat, freeze-thaw cycles, microwaves or sonication to lyse cells; followed by purification using spin columns, magnetic beads or the chemical properties of DNA to purify it via a

series of steps exploiting its differential solubility to other components within a sample. It is now very common to use specifically manufactured commercial kits which contain all reagents needed to extract DNA from a sample.

It is integral to any study reliant on 16S rRNA sequencing that the extraction method must provide an accurate depiction of the microbial composition in each sample. It is important to note that testing the DNA in a sample directly introduces far less bias than traditional culture methods, as well as permitting quantification of organisms and investigation of viable but not culturable organisms.

1.6. DNA extraction from *Staphylococcus aureus*

Staphylococcus aureus is a Gram-positive organism, the cell wall of Gram-positive species consists of a cytoplasmic cell membrane surrounded by a thick peptidoglycan layer (figure 2). This thick layer of peptidoglycan makes Gram-positive organisms significantly more difficult to lyse [34]. Even compared to other Gram-positive bacteria, *S. aureus* has a complex cell wall structure (Figure 3) due to its highly cross-linked peptidoglycan structure containing cross-linking peptides and teichoic acids which improve the rigidity of the cell wall [35]. Many *S. aureus* strains form a polysaccharide capsule which reinforces cellular integrity, making the organism even more robust [36]. Furthermore, *S. aureus* commonly forms complex multi-layered biofilms, leading to members of a community being physically protected by other members as well as an extracellular matrix which benefits the whole community [37]. The combination of these factors results in *S. aureus* being a highly robust organism and a difficult target for DNA extraction. Therefore, it is an ideal candidate for this study, as if *S. aureus* in the clinical samples is successfully lysed, it can be assumed that all other organisms will have also been lysed, giving an accurate representation of the lower respiratory tract microbiome.

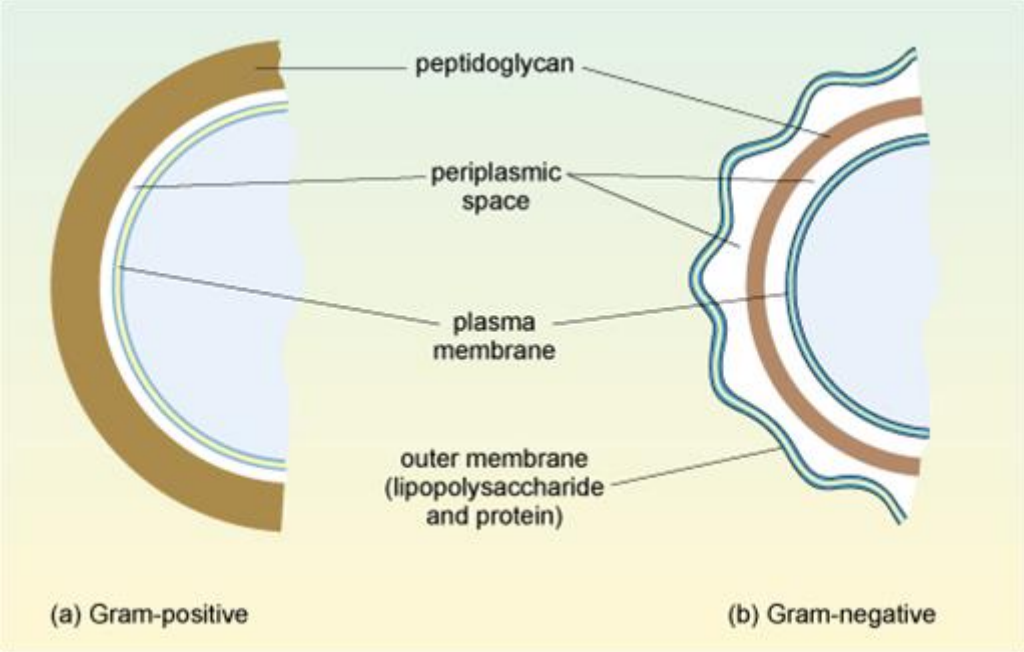


Figure 2 *- Basic cell wall structures of Gram-positive and Gram-negative bacteria, showing thick peptidoglycan layer which makes Gram positive organisms more resistant to cellular lysis (Source: The Open University, 2019)[38].*

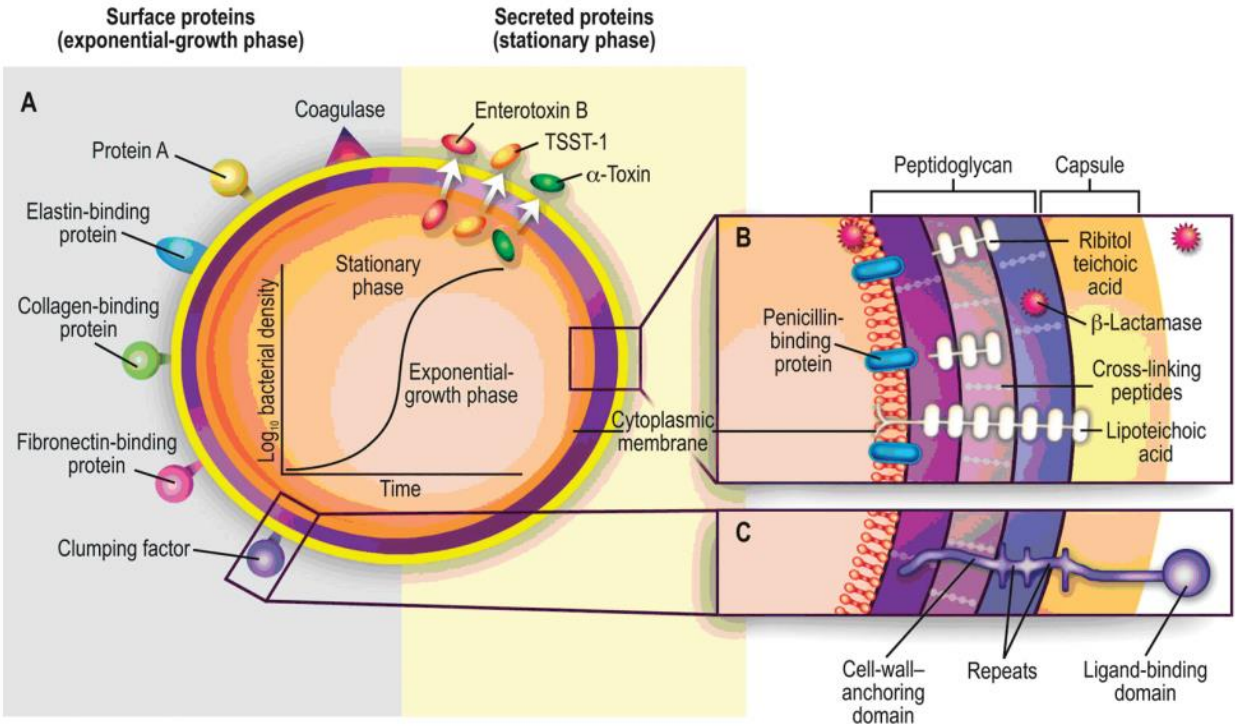


Figure 3 *- Staphylococcus aureus cell wall structure showing cross-linking in peptidoglycan layer. A shows surface and secreted proteins, B and C show cross-sections of the cell envelope, including cross-linking peptides and teichoic acids present in the peptidoglycan layer (Source: Gordon and Lowy, 2008)[39].*

2. Aims

The primary intention of this project was to develop and optimise a protocol for the extraction of microbial DNA from complex pulmonary clinical samples. *Staphylococcus aureus* was used as a model organism for this project as it is an important pulmonary pathogen and is a particularly difficult organism to lyse; therefore, if the protocol is successful when used on *S. aureus*, it can be expected to work with a very wide range of microorganisms. The protocol resulting from this project will be beneficial if used in the academic setting by advancing the techniques used for microbiome research, and will also be beneficial if used in the healthcare setting by improving the efficiency of pulmonary infection diagnoses compared to traditional culture based methods. To accomplish this main objective, several specific goals must be achieved by answering the questions below.

Selection of an extraction method:

1. Is there a difference in extraction ability when implementing a wide-spectrum multi-lytic enzyme mix (such as MetaPolyzyme) for use with a commercial DNA extraction kit compared to using a species-specific lysis enzyme (such as Lysostaphin)?
2. Is there a difference in yield when extracting DNA from fresh samples compared to frozen samples?
3. Is a commercial DNA extraction kit more effective than traditional phenol/chloroform methods at extracting low-yield microbial DNA?
4. Is enzymatic lysis more effective than physical/detergent based lysis methods for extracting low-yield microbial DNA?
5. Which of the candidate extraction methods is most effective at extracting DNA from low-yield microbial samples?

Development and optimisation of chosen method:

6. How reliable is this DNA extraction method?
7. Can application of Phase Lock Gel reduce the variability in results produced when using this DNA extraction method?
8. Is the limit of detection (LoD) of this DNA extraction method on pulmonary samples at least comparable to the current clinical cut off for a positive BAL sample ($1x10^4$ CFU/ml)?
9. Does the use of a complex clinical sample impair the efficacy of this DNA extraction method and its associated qPCR analysis?
10. Is it possible to perform absolute quantification of the number of cells present in the initial sample when using this DNA extraction method and its associated qPCR analysis?

3. Materials and methods

3.1. Reagents

Item	Manufacturer
325 mesh silicon dioxide	Sigma
70% ethanol	Sigma
Aluminium ammonium sulphate dodecahydrate	Acros Organics
Chloroform:Isoamyl alcohol (24:1)	Sigma
CTAB (Hexadecyltrimethylammonium bromide)	Sigma
CutSmart restriction enzyme buffer	New England Biolabs
Dimethylsiloxane-350	Mistral
EcoRI High-Fidelity restriction enzyme	New England Biolabs
GelRed nucleic acid gel stain	Biotium
GeneJET plasmid miniprep kit	Thermo Scientific
GeneJET plasmid miniprep kit	Thermo Scientific
GeneRuler 1kb	Thermo Scientific
GeneRuler 50bp	Thermo Scientific
Genomic DNA clean & concentrator kit	Zymo Research
GoTaq Green PCR master mix	Promega
High-pure PCR template preparation kit	Roche
KAPA HiFi HotStart ReadyMix	Roche
KAPA SYBR FAST qPCR master mix	KAPA Biosystems
KpnI High-Fidelity restriction enzyme	New England Biolabs
Linear polyacrylamide (LPA)	Sigma
Low-EDTA TE buffer	Made in-house
Lysing Matrix E tubes	MP Biomedicals
Lysostaphin	Sigma
MetaPolyzyme multilytic enzyme mix	Sigma
pCDF-Duet-1 plasmid vector	Novagen
PCR grade water	Promega
Phenol:Chloroform:Isoamyl alcohol (24:25:1)	Sigma
Phosphate buffered saline (PBS)	Made in-house
Polyethylene glycol 6000 (PEG)	Sigma
QIAquick PCR purification kit	Qiagen
qPCR grade water	Invitrogen
SYBR Safe DNA gel stain	Invitrogen
T4 DNA ligase	New England Biolabs
T4 DNA ligase reaction buffer with 10mM ATP	New England Biolabs

Item	Manufacturer
Tris-EDTA (TE) buffer	Made in-house
Tryptone soy agar (TSA)	Made in-house
Tryptone soy broth (TSB)	Made in-house

***Table 1** – Reagents used and suppliers. Reagents denoted as 'made in-house' were produced by the University of Bristol CMM laboratory technicians.*

3.2. Equipment

Equipment	Manufacturer
Centrifuge	VWR Microstar 17R
Electrophoresis powerpac	Bio Rad – powerpac 300
GeneSys UV transilluminator	SYNGENE – G:box
Heat block	Techne BRI block DB.2P
NanoDrop	Thermo Scientific - NanoDrop Lite Spectrophotometer
PCR Thermocycler	PCR Max alpha cycler
qPCR Thermocycler	Bio molecular systems – Magnetic induction cycler (mic)
Ribolyser (Bead-beater)	Fast prep hybaid fp120 fp120hy-230
Shaking incubator	New Brunswick Scientific - Innova 44
Static incubator	Leec compact incubator

***Table 2** - Equipment used and suppliers.*

3.3. Strains and growth conditions

For experiments comparing DNA extraction methods, the bacterial strain used was *Staphylococcus aureus* USA300 JE2, taken from a collection owned by Professor Ruth Massey (University of Bristol) and cultured using tryptone soy agar (TSA) and tryptone soy broth (TSB). For transformation of the *Vibrio* 16S rRNA gene in a TOPO TA pCR2.1 plasmid and the production of the p(*nuc*) standard, chemically competent DH5α *Escherichia coli* cells were purchased from Invitrogen and cultured using Luria-Bertani agar (LB agar) and Luria-Bertani broth (LB broth).

Strains were streaked onto agar plates and incubated at 37°C in a static incubator for 18 hours. A single colony was then taken from the agar plate and sub-cultured into the appropriate culture broth, briefly mixed and then incubated at 37°C in a shaking incubator for 18 hours.

3.4. DNA extraction methods

DNA extraction protocols used:

- Roche High Pure PCR Template Preparation Kit (Met+Kit)
- Bead-beating with phenol/chloroform purification (BB+P/C)
- MetaPolyzyme with phenol/chloroform purification (Met+P/C)

3.4.1. Roche High Pure PCR Template Preparation Kit

("Met+Kit")

DNA was extracted from bacterial cells in liquid culture following the manufacturers protocol with some minor deviations to improve yield and accuracy for the specific sample types used. 500µl of the sample was added to a 1.5ml microcentrifuge tube and centrifuged for 5 minutes at 3,000 x g, the pellet was then resuspended in 200µl sterile PBS. To lyse the cells in the sample, 10µl of MetaPolyzyme was added (10mg/ml) and incubated for either 15 minutes or 1 hour at 37°C (in a static incubator). Following incubation, 200µl of binding buffer and 100µl isopropanol was added and the sample was mixed well before being centrifuged for 5 minutes at 13,000 x g. The supernatant was transferred to a High Pure filter tube, fitted to a collection tube and centrifuged for 1 minute at 8,000 x g. At this point, the DNA in the sample has been removed from the cells and bound to glass fibres within the filter tube. The following steps are to remove contaminants and inhibitors before eluting the purified DNA.

After centrifugation, flow through liquid was discarded and the filter tube fitted to a new collection tube. 500µl inhibitor removal buffer was added before centrifuging for 1 minute at 8,000 x g. After centrifugation, flow through liquid was again discarded and a new collection tube fitted before adding 500µl wash buffer and centrifuging again for 1 minute at 8,000 x g. This wash step was repeated once, then after flow through had been discarded, centrifuged for 10 seconds at full speed to remove residual wash buffer. The collection tube was discarded, and the filter tube inserted into a clean, sterile 1.5ml microcentrifuge tube. 200µl prewarmed (70°C) elution buffer was added and left for 30 minutes at room temperature before being centrifuged for 1 minute at 8,000 x g to transfer the purified DNA from the filter tube into the microcentrifuge tube.

3.4.2. Bead-beating with phenol/chloroform purification *("BB+P/C")*

This protocol was kindly provided by Dr Michael Cox (University of Birmingham). DNA was extracted from either bacterial suspension or clinical samples. Sample preparation differs in this method depending on sample type. For bacterial isolates, the organism was grown in a 37°C shaking incubator for 18 hours in appropriate media, 2ml of this bacterial suspension was then centrifuged for 10 minutes at 16,000 x g. For bronchoalveolar lavage (BAL) or pleural fluid (PF), 5ml was centrifuged for 20 minutes at full speed.

For all sample types, the pellet was then resuspended in 500µl CTAB buffer and incubated for 15 minutes at room temperature. Each sample was then transferred into a Lysing Matrix E (LME) tube and 50µl aluminium ammonium sulphate (AAS) was added, followed by 500µl phenol:chloroform:isoamyl alcohol (PCI). The LME tubes were inserted into a ribolyser and run at a speed of 5.5m/sec for 1 minute (2 x 30 second runs). The LME tubes were then centrifuged for 5 minutes at 16,000 x g. All liquid was then transferred to a fresh 1.5ml microcentrifuge tube (containing 100µl phase lock gel for experiments using PLG) and the LME tube was left on ice. The microcentrifuge tubes were centrifuged at 4°C for 5 minutes at 16,000 x g, 500µl of chloroform:isoamyl alcohol (CI) added, briefly mixed and then centrifuged again at 4°C for 5 minutes at 16,000 x g. 500µl CTAB buffer, 50µl AAS and 500µl PCI was added to each LME tube which had been left on ice, inserted into the ribolyser and the extraction repeated.

After centrifugation with CI, the aqueous phase from each microcentrifuge tube was transferred to pre-prepared fresh 2ml microcentrifuge tubes containing 1µl Linear Polyacrylamide (LPA), 1ml of PEG/NaCl added and mixed well. Samples were left at 4°C for 18 hours overnight to precipitate, then centrifuged at 4°C for 20 minutes at 16,000 x g. PEG/NaCl was aspirated from the pellets, and the pellets then washed with 500µl ice-cold 70% ethanol and centrifuged for 5 minutes at 16,000 x g. This wash was repeated twice using 200µl ice-cold 70% ethanol, and the pellets air dried for 5 minutes before being resuspended in 30µl low-EDTA TE buffer. Resuspended total nucleic acid from each sample (primary and secondary extraction) was then combined into a single tube as a single 60µl DNA extract per sample.

3.4.3. MetaPolyzyme with phenol/chloroform purification

("Met+P/C")

This method was adapted from the BB+P/C method. The key difference between the two methods is the technique used for lysis of bacterial cells within the sample.

Samples were prepared as above, however once resuspended in 500µl CTAB extraction buffer, 10µl of MetaPolyzyme (10mg/ml) was added and the sample incubated at 37°C for 1 hour (instead of 15 minutes at room temperature. The remainder of the extraction remained the same, except for the use of 1.5ml microcentrifuge tubes in place of the LME tubes and the exclusion of the bead-beating step. Mechanical lysis was not required as MetaPolyzyme causes enzymatic lysis of the bacterial cells. This method also excluded the second extraction as the enzymes in the MetaPolyzyme mix would be degraded by the phenol already in the sample, rendering a secondary lysis stage ineffective.

3.5. DNA quantification and storage

DNA from all extractions was quantified using a NanoDrop spectrophotometer (Thermo Scientific) with the appropriate elution buffer used as a blank. The DNA was then stored at -20°C until required.

3.6. PCR

3.6.1. Primers

Primers were purchased from Eurofins Genomics as lyophilised powder. They were diluted to 100pmol/µl as directed on the oligonucleotide synthesis report provided by Eurofins to produce a stock; from this stock they were then diluted to a working concentration of 10pmol/µl. Forward and reverse primers for each primer pair were combined. Primers used can be found in Table 3 and IUPAC degeneracy codes used in the 16S rRNA primer set can be found in Table 4. The optimal annealing temperatures (T_a) for the primers used in traditional PCR were determined by temperature gradient PCR runs (shown in Appendix 9.1).

Primer	Oligonucleotide Sequence
16S rRNA V4 – 520 (F)	5'- AYTGGGYDTAAAGNG -3’
16S rRNA V4 – 802 (R)	5'- TACNVGGGTATCTAATCC -3’
*nuc*_Hoegh (F)	5′- GGGTTGATACGCCAGAAACG -3′
*nuc*_Hoegh (R)	5′- TGATGCTTCTTTGCCAAATGG -3′
*nuc*_qPCR_F	5’- ATTGAAGTCGAGTTTGACAAAG -3’
*nuc*_qPCR_R	5’- TTGTGCTTCACTTTTTCTTAAAAG -3’
RD_*nuc*_F_(EcoRI)	5'- ATAT-GAATTC-CTAAAAAGAAAGAGGTGTTAGTTATGAC -3’
RD_*nuc*_R_(KpnI)	5'- ATAT-GGTACC-GACACTTTTACAATGAGCATTATTG -3’
ACYCDuetUP1	5’- GGATCTCGACGCTCTCCCT -3’

Table 3 *- Primers used for PCR, qPCR and restriction digest procedures. All primers were synthesised by Eurofins genomics. Sources: the sequences for the 16S rRNA V4 primers were provided by Dr Michael Cox (University of Birmingham); the sequences for the nuc_Hoegh primers were taken from literature (Hoegh et al., 2014)[40]; the nuc_qPCR and RD_nuc primer sets were designed specifically for this project; and the ACYCDuetUP1 primer was designed by Novagen (Merck Group) for pCDF-Duet-1 bacterial vector [41].*

IUPAC nucleotide code	Corresponding Base
A	Adenine
C	Cytosine
G	Guanine
T (or U)	Thymine (or Uracil)
R	A or G
Y	C or T
S	G or C
W	A or T
K	G or T
M	A or C
B	C or G or T
D	A or G or T
H	A or C or T
V	A or C or G
N	Any base

Table 4 – *Nucleotide degeneracy codes from the International Union of Pure and Applied Chemistry (IUPAC)[42]. These codes were used in the sequence of the 16S rRNA primer set.*

3.6.2. PCR conditions

PCR reactions had a total volume of 25µl, comprised of 12.5µl GoTaq MasterMix (Promega), 2.5µl forward and reverse primer mix (1:1, each at 10pmol/µl), 2µl template DNA and 8µl PCR grade water (VWR). A PCR Max alpha cycler was used to perform the reactions, the conditions used unless otherwise stated can be found in Table 5.

Step	Temperature (°C)	Time (seconds)	Cycles
Initial denaturation/polymerase activation	95	120	1
Denaturation Annealing Extension	95 (39.5 or 58.2)* 72	30 30 60	30
Final Extension	72	300	1
Final Store	4	∞	-

Table 5 *- PCR reaction conditions used. *Annealing temperatures used differed depending on the primer set used: 39.5°C for 16S rRNA primers, 58.2°C for nuc_Hoegh primers.*

3.6.3. Agarose gel electrophoresis

The PCR products were visualised using electrophoresis on a 1.5% agarose gel. The gel was prepared dissolving 1.5g agarose in 100ml 1x TE buffer and then poured into a cradle with SYBR Safe gel stain added at 0.1µl/ml. 10µl of each sample was loaded alongside 10µl of Thermo Scientific GeneRuler DNA ladder (either 50bp or 1kb) and gels were run at 90V, 300mA for 60 minutes using a Bio-Rad Power pac 300. Gels were visualised and photographed using a GeneSys UV transilluminator.

3.6.4. Quantitative PCR (qPCR)

qPCR reactions had a total volume of 20µl, comprised of 10µl KAPA SYBR FAST master mix, 1µl of mixed forward and reverse (1:1) primers, 2µl of DNA and 7µl of qPCR grade water (Invitrogen). Reactions were run on a magnetic induction cycler (mic); conditions shown in Table 6.

Step	Temperature (°C)	Time (seconds)	Cycles
Initial denaturation/polymerase activation	95	180	1
Denaturation Annealing Extension	95 (50 or 60)* 72	20 30 30	40
Melt conditions (default)	50 to 80	100 (0.3°C/s)	1

***Table 6** – qPCR reaction conditions used. *Annealing temperatures used differed depending on the primer set used: 50°C for 16S rRNA primers, 60°C for nuc_qPCR primers.*

3.7. Plasmid standards

3.7.1. Preparation of the p(16S) standard

An aliquot of the p(16S) standard used was kindly donated by Dr Michael Cox (University of Birmingham). The plasmid standard used was a TOPO TA pCR2.1 vector containing the full length 16S rRNA gene from the *Vibrio natriegens* strain DSMZ 759. It was used at a working concentration of $2x10^7$ copies/µl. A plasmid map of the *V. natriegens* 16S rRNA gene in a TOPO TA pCR2.1 vector can be found in Appendix 9.2.

The plasmid was transformed into DH5α *E. coli* cells by heat shock. A 50µl aliquot of chemically competent DH5α cells were taken from the -80°C freezer and defrosted on ice. Once defrosted, 5µl of the original p(16S) standard was added and left on ice for 30 minutes. The sample was then heat shocked in a 42°C water bath for 2 minutes, before being placed on ice for a further 5 minutes. 750µl LB broth was then added, mixed briefly and then incubated at 37°C in a shaking incubator for an hour. After incubation, the sample was centrifuged for 1 minute at 12,000 x g. Most of the supernatant was removed, leaving ~100µl in which the pellet was then resuspended before being streaked onto a plate of LB agar containing ampicillin (100µg/ml) and incubated at 37°C for 18 hours. After incubation, single colonies were taken, added to 5ml of LB broth and incubated at 37°C for 18 hours. A stock of DH5α containing the p(16S) plasmid was produced by diluting this culture at a 1:1 ratio with 50% glycerol and frozen at -80°C.

To produce a working stock, some DH5α cells containing the p(16S) plasmid were taken from the stock at -80°C and grown in LB broth at 37°C for 18 hours. The plasmid was recovered from this culture using a GeneJET Plasmid Miniprep Kit according to the manufacturer's instructions. The

recovered plasmid was quantified using a NanoDrop spectrophotometer. To dilute the recovered plasmid to a working concentration of $2x10^7$, the total length of the plasmid was calculated based on the plasmid length stated in the TOPO TA pCR2.1 user guide [43] (3931 base pairs) and the length of the inserted *Vibrio natriegens* DSMZ 759 16S rRNA gene (1465 base pairs). Copy number per µl was calculated using the online URI Genomics & Sequencing Center copy number calculator [44] and the sample diluted to give a final concentration of $2x10^7$.

3.7.2. Production of the p(*nuc*) standard

The p(*nuc*) plasmid standard produced was a pCDF-Duet-1 vector containing the full length *nuc* gene from the *Staphylococcus aureus* strain JE2. The p(*nuc*) standard was used at a working concentration of $2x10^7$ copies/µl. A plasmid map of the *S. aureus nuc* gene in a pCDF-Duet-1 vector can be found in Appendix 9.3.

The first stage in the production of a p(*nuc*) standard was to design suitable primers for the restriction digest of the *nuc* gene out of the *S. aureus* genome. Both primers were designed to contain a restriction site at the 5' end and an ATAT repeat was also added at the 5' end of each restriction site to increase the binding capability of the restriction enzymes to the sequences during the restriction digest (Figure 4). An EcoRI site was added to the forward primer and a KpnI site was added to the reverse primer. The total *nuc* amplicon length was 749bp including the primers, restriction sites and ATAT repeats, and can be found in Appendix 9.4.

Restriction digest primer design:	5'– ATAT repeat –	Restriction sequence –	Primer sequence – 3'
RD_nuc_F_(EcoRI):	5'– ATAT –	GAATTC –	CTAAAAAGAAAGAGGTGTTAGTTATGAC – 3'
RD_nuc_R_(KpnI):	5'– ATAT –	GGTACC –	GACACTTTTACAATGAGCATTATTG – 3'

Figure 4 *- Restriction digest primer design. Both primers were designed to contain a restriction site at the 5' end, with an ATAT repeat to assist restriction enzyme binding to DNA.*

The vector used for the p(*nuc*) standard was pCDF-Duet-1. This vector was selected as it was not possible to use the TOPO-pCR-2.1 vector used for the p(16S) plasmid, and pCDF-Duet-1 is similar in size (TOPO-pCR-2.1 = 3931bp; pCDF-Duet-1 = 3781bp). The pCDF-Duet-1 plasmid used was kindly donated by John Shaw (University of Bristol).

To isolate the *nuc* fragment, 1µl of genomic DNA extracted from the *S. aureus* strain JE2 using the Roche High Pure PCR Template Preparation Kit was used according to the manufacturer's instructions. The 1µl of JE2 DNA was amplified by PCR using the RD_nuc_F_(EcoRI) and RD_nuc_R_(KpnI) primers. The PCR reaction had a total volume of 20µl, comprised of 10µl Kapa Hifi HotStart ReadyMix, 1µl of each primer, 1µl of the JE2 genomic DNA template and 7µl of PCR grade water. Two samples were run, along with a negative control which substituted the DNA template for PCR grade water. The conditions of this PCR reaction can be found in Table 7.

Step	**Temperature (°C)**	**Time (seconds)**	**Cycles**
Initial denaturation/ polymerase activation	95°C	300 (5 min)	1
Denaturation Annealing Extension	95°C 60°C 72°C	30 30 60	35
Final Extension	72°C	300 (5 min)	1
Final Store	10°C	∞	-

***Table 7** – PCR conditions used for isolation of nuc gene fragment during cloning procedure to produce p(nuc) plasmid standard.*

To check the amplified *nuc* fragment produced by PCR, the PCR products were run on a 1.5% agarose gel containing 0.1µl/ml SYBR-Safe gel stain. 5µl of each PCR product was loaded into respective wells and electrophoresis was carried out at 300mA, 90V for 30 minutes. This confirmed a fragment of the expected size in the template positive samples, which was absent in the negative sample.

The remaining PCR product for the *nuc* positive samples was purified using a Zymo Genomic DNA clean and concentrator kit according to the manufacturer's instructions. The product was eluted in 15µl of the elution buffer provided in the kit, and DNA concentration of both samples quantified

using a NanoDrop Spectrophotometer (207.7 ng/µl and 211.1 ng/µl respectively). The purified *nuc* fragments were stored at -20° until required.

To prepare the plasmid vector, 1µl of pCDF-Duet-1 purified plasmid (at 172.1 ng/µl) was added to 50µl of competent DH5α cells in a microcentrifuge tube and mixed by gentle pipetting. The tubes were then left on ice for 30 minutes before being subjected to heat shock at 42°C for 2 minutes on a thermal block. After the heat shock, the tubes were placed back on ice for 3 minutes, then 700µl LB broth was added and the tubes were incubated at 37°C for 1 hour in a shaking incubator. After incubation, the tubes were centrifuged at 8,000 RPM for 5 minutes, the majority of the supernatant discarded and the pellet resuspended in the remaining ~100µl of supernatant before being plated out on LB agar plates containing 30µg/ml streptomycin to select for DH5α cells containing the pCDF-Duet-1 plasmid. Plates were incubated at 37°C for 18 hours, then a single colony was transferred to 5ml LB broth containing 30µg/ml streptomycin and incubated at 37°C for a further 18 hours in a shaking incubator. After incubation, two 900µl aliquots were combined 1:1 with 50% glycerol and stored at -80°C. The plasmid was reclaimed from the remaining DH5α cells using a Thermo Scientific GeneJet plasmid miniprep kit according to the manufacturer's instructions. The plasmid was eluted in 50µl PCR grade water and quantified using a NanoDrop Spectrophotometer (209.5 ng/µl).

The pCDF-Duet-1 plasmid vector and *nuc* fragment were then cut via restriction digestion using EcoRI and KpnI restriction enzymes (Figure 5). First, the purified plasmid and *nuc* fragment were both diluted to 200ng/µl. Next, 10µl (2µg) of either pCDF-Duet-1 or the *nuc* fragment was combined with 1µl EcoRI-HF, 1µl KpnI-HF, 5µl CutSmart buffer and 33µl PCR grade water. As a control, two single digests were also carried out on pCDF-Duet-1 using EcoRI and KpnI respectively. Reagents were combined, mixed thoroughly and incubated at 37°C for 4 hours. After incubation, 5µl of each restriction digest product was run on a 1% agarose gel containing 0.04µl/ml Gel-Red stain for 1 hour at 300mA and 90V. 3µl of 1kb ladder, 3µl of 50bp ladder and 3µl of uncut pCDF-Duet-1 was also run for comparison. The remaining restriction digestion products were purified using a QIAquick PCR purification kit according to the manufacturer's instructions.

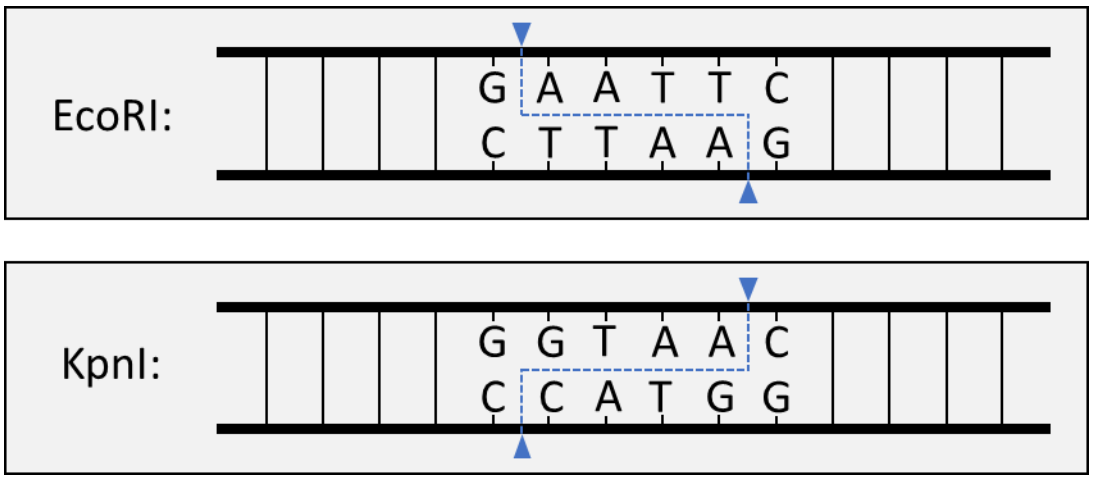


Figure 5 *– Diagram of endonuclease activity of EcoRI and KpnI restriction enzymes at their sequence specific restriction sites.*

To ligate the cut *nuc* fragment into the cut pCDF-Duet-1 plasmid, the insert and plasmid were combined at a 3:1 (insert:plasmid) ratio. Following the restriction digest, the cut *nuc* fragment was 699bp in length and at a concentration of 31.2ng/µl; the cut pCDF-Duet-1 plasmid was 3551bp in length and at a concentration of 25.0ng/µl. The ligation reaction contained 100ng of DNA and had a total volume of 10µl; this was comprised of 4µl of the cut pCDF-Duet-1 plasmid, 1.9µl of the cut *nuc* fragment, 1µl of NEB 10x buffer for T4 DNA ligase with 10mM ATP, 1µl of NEB T4 DNA ligase, and 2.1µl of PCR grade water. An insert negative reaction was also performed which substituted the 1.9µl of cut *nuc* fragment for PCR grade water. All reagents were combined in a PCR tube and left for 2 hours at room temperature. The result was a pCDF-Duet-1 plasmid vector containing the *nuc* insert, henceforth designated as p(*nuc*). The p(*nuc*) plasmid was transformed into DH5α cells via heat shock using the same method stated above for the transformation of pCDF-Duet-1 into DH5α cells; a stock was combined 1:1 with 50% glycerol and stored at -80°C.

Successful ligation of the *nuc* insert into pCDF-Duet-1, and subsequent transformation of p(*nuc*) into DH5α cells was confirmed by colony PCR. 10 single colonies of DH5α containing p(*nuc*), and 2 colonies of DH5α not containing the p(*nuc*) plasmid, were added to independent microcentrifuge tubes containing 20µl of PCR grade water and mixed by pipetting. For each colony mix, 3µl was added to a PCR tube and combined with 10µl GoTaq Green Mastermix, 1µl of each primer and 5µl of PCR grade water. The primers used were ACYCDuetUP1 which is specific to a sequence on the pCDF-Duet-1 plasmid, and RD_*nuc*_R_(KpnI) which is specific to the *nuc* insert. This ensured that

logarithmic amplification would only occur if the DH5α colony contained pCDF-Duet-1 with the *nuc* fragment inserted (the p(*nuc*) plasmid). The conditions of the PCR reaction used can be found in Table 8. To visualise the results by electrophoresis, 7.5µl of each PCR product was run on a 1% agarose gel containing 0.1µl/ml of SYBR-Safe gel stain against 3µl of a 1kb DNA ladder and 3µl of a 50bp DNA ladder. All 10 colonies of DH5α containing p(*nuc*) produced positive results, and both colonies of DH5α not containing the p(*nuc*) plasmid produced negative results.

Step	**Temperature (°C)**	**Time (seconds)**	**Cycles**
Initial denaturation/ polymerase activation	95°C	900 (15 min)	1
Denaturation Annealing Extension	95°C 60°C 72°C	30 30 120	35
Final Extension	72°C	300 (5 min)	1
Final Store	10°C	∞	-

***Table 8** – Conditions used for colony PCR to check presence of p(nuc) standard in DH5α cells following transformation by heat shock.*

To prepare the p(*nuc*) standard, the plasmid was extracted from the p(*nuc*) positive DH5α cells using a GeneJET plasmid miniprep kit according to the manufacturer's instructions. The purified plasmid was eluted in 50µl of the elution buffer provided in the kit and quantified using a NanoDrop Spectrophotometer; the concentration of the p(*nuc*) plasmid was 199.1ng/µl. The purified plasmid was checked for the *nuc* insert by restriction digest using EcoRI and KpnI as above. Once confirmed, the remaining plasmid was diluted to $2x10^7$ copies/µl. The p(*nuc*) plasmid was 4280bp in length and at a concentration of 199.1ng/µl; copy number per µl was determined as $4.31x10^{10}$ using the online URI Genomics & Sequencing Center copy number calculator [44]. To dilute the plasmid to $2x10^7$ copies/µl, 1µl of p(*nuc*) was added to 2.155ml of qPCR grade water. The p(*nuc*) standard at a concentration of $2x10^7$ copies/µl was stored at -20°C until required.

3.8. Phase lock gel

3.8.1. Production of phase lock gel

Phase lock gel (PLG) was made by combining 850g of dimethylsiloxane-350 with 150g of 325-mesh silicon dioxide, mixing well and leaving in a 37°C shaking incubator for a week to mix fully. Once mixed, the gel was sterilised by autoclaving at 121°C for 15 minutes. Directions for making PLG were found on the BiteSizeBio and PipetteJockey websites [45, 46]. This produced a gel equivalent to commercial 'Phase Lock Gel: Heavy' [47].

3.8.2. Phase lock gel standardisation experiment

This experiment aimed to determine the ability of phase lock gel (PLG) to reduce the variability between results for identical samples processed using the BB+P/C DNA extraction procedure. Three biological repeats were performed, with three technical repeats per biological repeat.

For each biological repeat, a 10ml liquid culture of *S. aureus* strain JE2 was produced as described previously, 6ml was taken to produce 6 aliquots of 1ml each; three of these were designated as the PLG-positive (PLG+) group, the remaining three were the PLG-negative (PLG –) group. 6 aliquots of 1ml TSB were also produced and three added to each group.

This resulted in four sample types:

- JE2 PLG +
- JE2 PLG –
- TSB PLG +
- TSB PLG –

All samples were subjected to the BB+P/C DNA extraction protocol in parallel. Following the extraction, DNA for each sample was resuspended in 60µl low-EDTA TE buffer and yield was quantified using a NanoDrop spectrophotometer. Three readings were taken per sample and an average taken to determine the yield.

3.9. Extended serial dilution experiments: Limit of detection (LoD)

These experiments aimed to determine the sensitivity of the BB+P/C DNA extraction and associated qPCR analysis for each primer set used. This was done by calculating the limit of detection (LoD) and independent experiments were carried out to determine the LoD of each primer set for:

1. The qPCR analysis stage in isolation
2. The BB+P/C DNA extraction and qPCR analysis stage combined

To determine the LoD of the qPCR stage in isolation both plasmid standards, p(*nuc*) and p(16S) (each at $2x10^7$ copies per µl), were combined with 5µl of qPCR grade water to dilute them to $1x10^7$ per µl. A seven-fold serial dilution was then performed by adding 1µl of each standard to 9µl of qPCR grade water in series to produce set of samples containing a range from $1x10^7$ copies per µl to 1 copy per µl. Two samples of 2µl were taken from each dilution and analysed by qPCR using the respective primer set (16S rRNA or *nuc*_qPCR). Each qPCR run also included two non-template control (NTC) samples containing 2µl qPCR grade water instead of the plasmid standard.

To determine the LoD of the BB+P/C DNA extraction and qPCR analysis stages combined, a 10ml liquid culture of *S. aureus* strain JE2 was produced as described previously. A thirteen-fold serial dilution was performed by taking 1ml from each JE2 containing dilution and adding it to 9ml of TSB, producing a series from the neat culture (N) to $1x10^{-14}$. This range was chosen as it was expected to pass beyond a dilution of 1 CFU/ml. For each dilution, three 1ml aliquots were added to individual petri dishes and TSA at 37°C was added, mixed briefly and left to set. Plates were incubated as described previously and then colonies counted and an average taken for each dilution. For each dilution containing at least 1 CFU/ml (N to $1x10^{-9}$), three individual 1ml samples were processed using the BB+P/C DNA extraction method along with three blank samples of 1ml TSB. DNA resuspended in 60µl low-EDTA TE buffer and 2µl from each DNA extract was analysed by qPCR with each set of primers (16S rRNA and *nuc*_qPCR).

3.10. Impact of pleural fluid on efficacy of BB+P/C DNA extraction and qPCR analysis

3.10.1. Pleural fluid

These experiments aimed to determine the impact of pleural fluid on the efficacy of the BB+P/C DNA extraction method. Clinical samples were collected by the Pleural Service at Southmead Hospital in Bristol as part of routine clinical care, ethical approval was through the North Bristol Pleural Database. The pleural fluid used in these experiments was obtained from a lymphoma patient who suffered pleural effusion, requiring pleurocentesis to remove the fluid via a pleural catheter. Excess fluid of ~65ml was collected and stored at -20°C. The following day it was transported on ice to the University of Bristol where it remained stored at -20°C until required. The pleural fluid was defrosted on ice for 45 minutes and 30ml was transferred to a fresh falcon tube and vortexed to homogenise the sample; this 30ml was then separated into six 5ml aliquots. Six 5ml aliquots of PBS were also produced for these experiments as a non-pleural fluid control.

3.10.2. BB+P/C and qPCR efficacy check – *S. aureus* spike in clinical samples

A 10ml liquid culture of *S. aureus* strain JE2 was produced as described previously and a 9-fold serial dilution was performed by taking 1ml from each JE2 containing dilution and transferring it to 9ml of TSB. 1ml from each dilution was plated out in triplicate and incubated as described previously. Following incubation the colonies were counted, an average CFU/ml was taken for each dilution and these averages were then used to estimate the CFU/ml of the neat liquid culture of JE2 (~$1.83x10^9$ CFU/ml). A 10µl spike of neat liquid culture (containing ~$1.83x10^7$ cells) was added to three of the six 5ml aliquots of pleural fluid and to three of the six 5ml aliquots of PBS to produce four sample types with three replicates each:

- Pleural fluid + ~$1.83x10^7$ cells of JE2
- Pleural fluid + TSB
- PBS + ~$1.83x10^7$ cells of JE2
- PBS + TSB

All twelve samples were then processed using the BB+P/C DNA extraction; following the extraction, the DNA was resuspended in 60ul low-EDTA TE buffer.

To determine the impact of the pleural fluid on the BB+P/C DNA extraction and qPCR stages combined, all DNA extracts were analysed by qPCR using both the 16S rRNA primers and the *nuc*_qPCR primers as described above. A standard curve of each respective plasmid standard (p(16S) or p(*nuc*)) at dilutions of $2x10^7$, $2x10^6$ and $2x10^5$ and a non-template control of qPCR grade water was also analysed in each qPCR run.

3.10.3. qPCR stage only efficacy check – pleural fluid spike in p(16S) and p(*nuc*) samples

To determine the impact of the pleural fluid on the qPCR analysis stage in isolation the plasmid standards, p(16S) and p(*nuc*), were analysed independently. The same method was used to analyse the impact of pleural fluid on qPCR analysis for each primer set. For each primer set, four samples containing 1µl of the relevant plasmid standard at $2x10^7$ copies per µl were produced. 1µl of DNA from the un-spiked pleural fluid sample was added to two of these samples, and 1µl of qPCR grade water was added to the remaining two tubes. This produced four sample types:

- Plasmid standard + pleural fluid derived DNA
- Plasmid standard + qPCR grade water
- qPCR grade water + pleural fluid derived DNA
- qPCR grade water only

All samples for each primer set were analysed by qPCR as described above and an average Cq value was taken from the two technical repeats per sample type.

4. Results 1 – Selection of DNA extraction method

4.1. Comparison of MetaPolyzyme vs. lysostaphin for cellular lysis when using a kit.

One of the DNA extraction methods used was the Roche High Pure PCR Template Preparation kit, these experiments were performed to determine if there is a difference in the DNA yield produced from identical samples of *S. aureus* when using MetaPolyzyme rather than lysostaphin. If there is no significant difference, MetaPolyzyme can replace lysostaphin for all future experiments in this project which use the Roche Kit.

The Roche Kit instructions recommends the use of lysozyme for isolation of DNA from bacteria, however, to isolate DNA from *S. aureus* cells the enzyme lysostaphin is routinely substituted for the recommended lysozyme, as this is needed to break down their thick peptidoglycan cell wall. The aim of this project is to be able to isolate DNA from any microorganisms present in a sample, not just *S. aureus,* so it was necessary to replace the highly specific lysostaphin enzyme with a broad-spectrum multi-lytic enzyme mix, such as MetaPolyzyme which contains mutanolysin, achromopeptidase, lyticase, chitinase, lysostaphin and lysozyme. As lysostaphin is routinely used, but MetaPolyzyme is more appropriate for the objectives of this project, it was logical to directly compare the ability of each enzyme when used for the cellular lysis step during the Roche Kit DNA extraction protocol. This experiment aimed to determine whether MetaPolyzyme could be used instead of lysostaphin to produce an equivalent yield of DNA from the kit-based method.

An overnight culture of *S. aureus* (JE2) was divided into 500µl samples and DNA was extracted from all samples using the Roche kit. Three biological repeats were performed, with three technical replicates per biological repeat. For each biological repeat, half of the samples were processed using lysostaphin (Lys+Kit) for cellular lysis, the other half were processed using MetaPolyzyme (Met+Kit). All samples were incubated at 37°C for 15 minutes after the addition of lytic enzymes. The extracted DNA was quantified using a NanoDrop spectrophotometer. Comparison of the amount of DNA extracted from each sample set showed no significant difference ($P=0.223$, two-tailed, unpaired t-test). Mean quantities are shown in Figure 6 and can be found in Appendix 9.5.

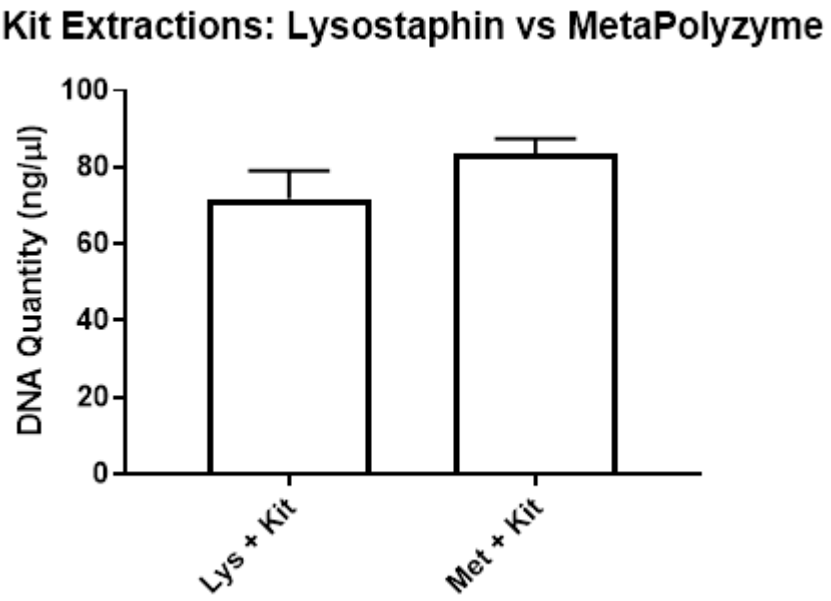


Figure 6 *– Mean quantities of DNA extracted from a JE2 culture using the Roche kit and either lysostaphin (Lys+Kit) or MetaPolyzyme (Met+Kit) for cellular lysis. Mean DNA quantity for Lys+Kit = 71.67ng/µl; mean DNA quantity for Met+Kit = 83.63ng/µl. DNA was quantified using a NanoDrop spectrophotometer. Error bars show standard error. A t-test showed no significant difference between groups: P = 0.223; T = 1.442; N = 6; df = 4. Raw data can be found in Appendix 9.5.*

To ensure that the *S. aureus* DNA extracted from these samples was of equivalent amounts, the extracts were amplified by PCR using the 16S rRNA primers to show an approximate quantity of total bacterial DNA and the *nuc*_Hoegh primers to show a rough quantity of staphylococcus specific DNA. The PCR product was then visualised by agarose gel electrophoresis (Figure 7).

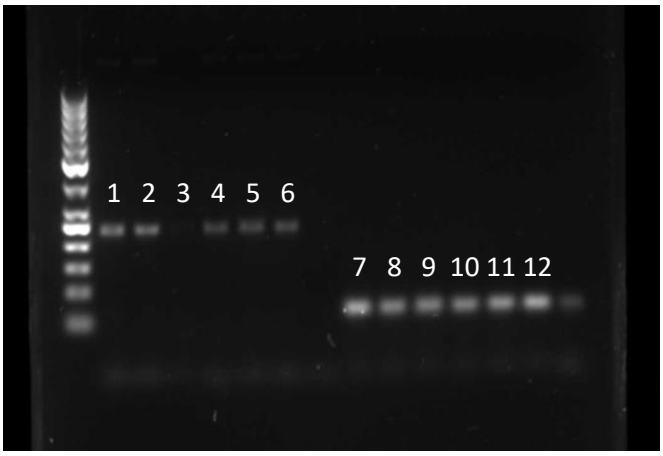


Figure 7 *– Agarose gel to visualise PCR product amplified using 16S rRNA primers (bands 1-6) and nuc_Hoegh primers (bands 7-12) after DNA extraction against a 50bp ladder. Bands 1-3 and 7-9 were from Met+Kit extractions; bands 4-6 and 10-12 were from Lys+Kit extractions. Image captured using GeneSys UV transilluminator.*

All bands for each primer set were of a similar brightness, again showing that the DNA extracted was of similar quantities. As there was no significant difference in the quantity of *S. aureus* DNA extracted when using MetaPolyzyme rather than lysostaphin, all further kit-based extractions were performed using MetaPolyzyme with confidence that any *S. aureus* in the samples would be successfully lysed.

4.2. Comparison of extraction ability with fresh vs. frozen samples

One consideration for studies using pulmonary clinical samples is the condition in which they will be provided. This is important as sample conditions prior to laboratory processing may significantly influence the results obtained [48]. The aim of this experiment was to determine if there is a significant difference in yield when extracting DNA from fresh or frozen samples as these are the two conditions in which clinical samples may be provided. Three DNA extraction methods were used: the Roche Kit with MetaPolyzyme method (Met+Kit), and two non-kit-based methods. One where MetaPolyzyme is used for cellular lysis followed by the use of phenol/chloroform for purification of the sample and ethanol precipitation of the DNA (Met+P/C); and one where bead-beating is used to mechanically lyse the cells, followed by phenol/chloroform purification and ethanol precipitation (BB+P/C).

Freezing a sample is a significant change in environmental conditions for the microorganisms within the sample, species react to freezing in different ways which may cause alterations in the relative viable microbial composition of the sample [49]. It is advantageous to be able to freeze samples for storage, but only if they can still provide an accurate depiction of the patient's microbiome; therefore, it is important to have a DNA extraction method which has as little difference between fresh and frozen samples as possible. All DNA extractions for each method were performed in parallel. In order to extract in parallel, frozen samples were from a different *S. aureus* (JE2) overnight culture to the fresh samples. Frozen samples were produced by dividing the overnight culture into 500µl samples and leaving at -20°C for 24 hours. For each extraction method, three fresh and three frozen biological repeats were processed, with three technical repeats for each biological replicate. Average DNA quantities from each method are shown in Figure 8 and can be found in Appendix 9.6.

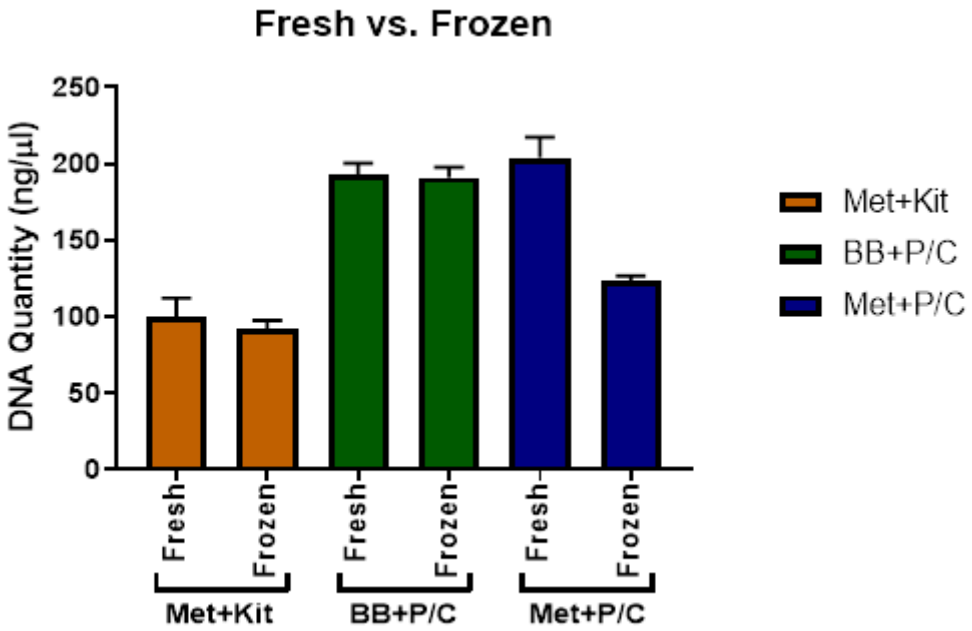


Figure 8 *– Average quantities of DNA for each extraction method on either fresh or frozen cultures of JE2: Fresh Met+kit = 99.38ng/µl; Frozen Met+kit = 91.78ng/µl; Fresh BB+P/C = 192.83ng/µl; Frozen BB+P/C = 191.03ng/µl; Fresh Met+P/C = 203.87ng/µl; Frozen Met+P/C = 123.14ng/µl. DNA was quantified using a NanoDrop spectrophotometer. Error bars show standard error. T-tests (fresh vs frozen): N = 6; df = 4; no significant difference between fresh and frozen for Met+Kit (P = 0.616; T = 0.543) or BB+P/C (P = 0.748; T = 0.345); very significant difference between fresh and frozen for Met+P/C (P = 0.005; T = 5.748). Raw data can be found in Appendix 9.6.*

The BB+P/C extraction method showed no significant difference between processing fresh and frozen samples (P=0.748, 2 tailed Students t-test), producing a relatively high yield in both sample types. The Met+Kit method also showed no significant difference between sample types (P=0.616, 2 tailed Students t-test) but produced a far lower yield than the BB+P/C method. The Met+P/C method showed the greatest difference between fresh and frozen samples (P=0.005, 2 tailed Students t-test), producing a relatively high yield from the fresh samples, and a relatively low yield from the frozen samples. These results suggest that the BB+P/C method may produce the most accurate results if the state of each sample cannot be controlled.

4.3. Comparison of kit vs. phenol/chloroform DNA extraction methods on neat culture

To examine the relative yields obtained by the kit and non-kit-based methods for the extraction of total genomic DNA from *S. aureus*, the Roche Kit with MetaPolyzyme method (Met+Kit) was directly compared with the two non-kit-based methods (Met+P/C and BB+P/C. Three biological repeats were performed, with three technical replicates for each biological repeat. All samples used were 500µl from a neat overnight culture of *S. aureus* (JE2) and for each biological repeat, all DNA extraction methods were performed on the same day in parallel. Extracted DNA was quantified using a NanoDrop spectrophotometer, mean quantities for each method are shown in Figure 9 and can be found in Appendix 9.7.

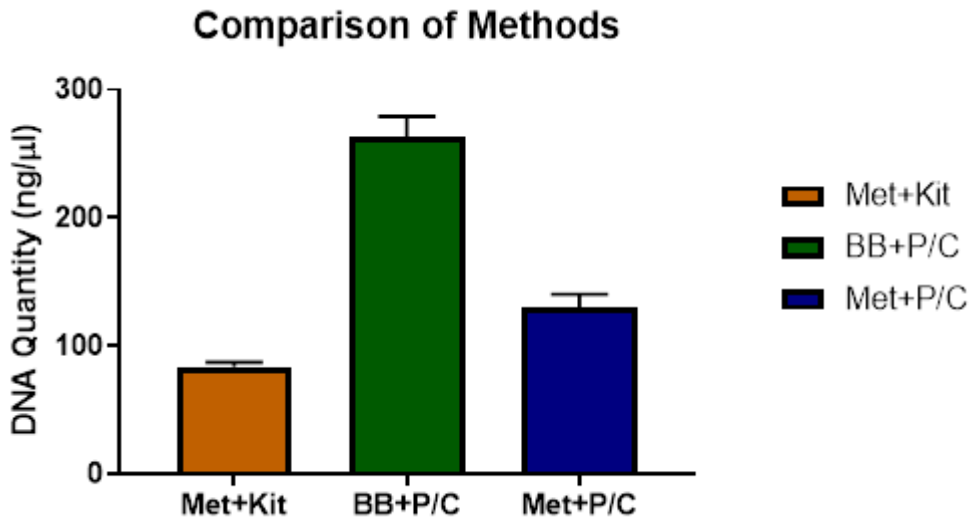


***Figure 9** – Mean quantities of DNA using each extraction method from the same culture of JE2. Mean DNA quantity for Met+kit = 83.63ng/µl; mean DNA quantity for BB+P/C = 262.73ng/µl; mean DNA quantity for Met+P/C = 130.01ng/µl. For each condition, 3 samples were measured in triplicate using a NanoDrop spectrophotometer. Error bars show standard error. Raw data can be found in Appendix 9.7.*

The BB+P/C method produced the highest yield, however also showed the greatest variance between samples. Both the Met+Kit and Met+P/C methods produced significantly lower yields from the same starting sample. However, there is a chance that the Met+Kit and Met+P/C methods were saturated by the relatively high number of cells in a neat sample and would be more effective at a lower dilution. It is important that the DNA extraction method is able to work effectively at lower dilutions as pulmonary samples typically have a low microbial load.

4.4. Comparison of DNA extraction methods on serially diluted culture

Up to this point, relatively high quantities of *S. aureus* cells were used for the extractions (~$1x10^8$-$1x10^9$ colony forming units (CFU)), but it is likely that the CFU in the clinical samples will be considerably lower. As the efficiency of the extraction methods may be affected by the abundance of the starting material, this experiment aimed to compare the sensitivity of each DNA extraction method on samples with lower bacterial loads. A serial dilution was performed on an overnight culture of *S. aureus* (JE2) from neat (N, between $1x10^8$ and $1x10^9$ CFU) to $1x10^{-2}$ (1 in 100 dilution). These dilutions were then split into 500µl samples and three biological repeats were performed per dilution for each extraction method. Three technical repeats were performed for each biological repeat. Extracted DNA was quantified using a NanoDrop spectrophotometer. Mean quantities for each dilution and method are shown in Figure 10 and can be found in Appendix 9.8.

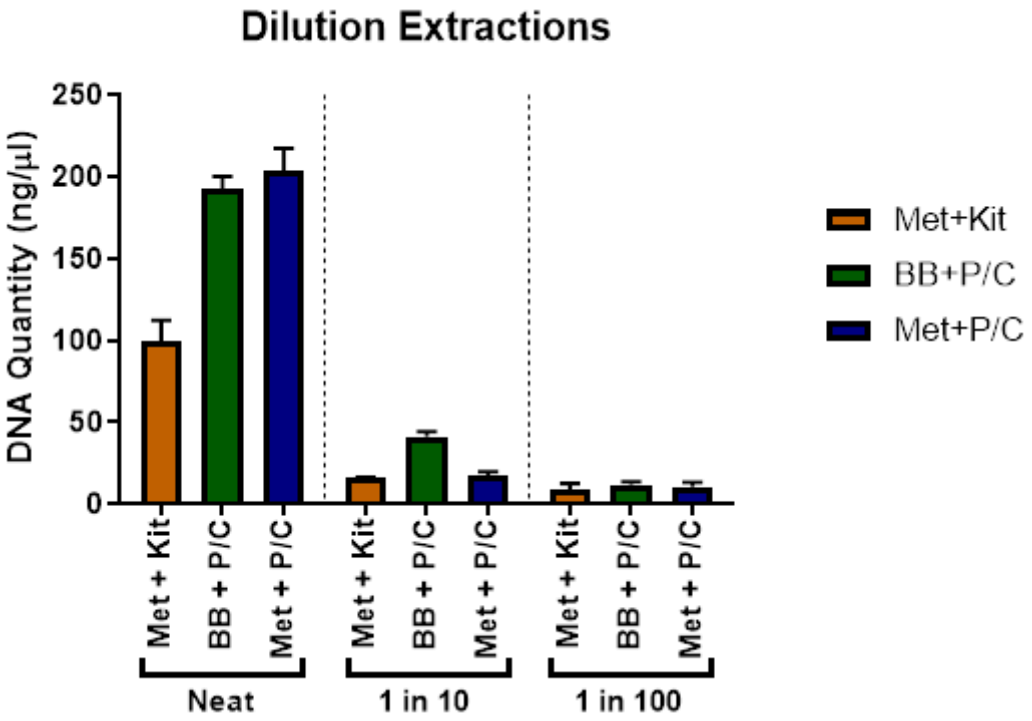


Figure 10 *– Mean quantities of DNA for each extraction method on samples at different dilutions (Neat, 1 in 10, and 1 in 100): Neat Met+kit = 99.38ng/µl; Neat BB+P/C = 192.83ng/µl; Neat Met+P/C = 203.87ng/µl; 1 in 10 Met+kit = 15.82ng/µl, 1 in 10 BB+P/C = 41.32ng/µl; 1 in 10 met+P/C = 17.17ng/µl; 1 in 100 Met+kit = 8.48ng/µl; 1 in 100 BB+P/C = 11.64ng/µl; 1 in 100 Met+P/C = 10.53ng/µl. For each condition, 3 samples were measured in triplicate using a NanoDrop spectrophotometer. Error bars show standard error. Raw data can be found in Appendix 9.8.*

In the neat samples, the Met+P/C method produced the highest yield (204ng/µl), followed by the BB+P/C method (193ng/µl), and the Met+Kit method produced the lowest yield 99ng/µl by a significant margin. The DNA quantities significantly decreased between the neat and $1x10^{-1}$ (1 in 10) dilution as expected, however the sensitivity of the NanoDrop was not high enough to provide an accurate quantification of the DNA at lower dilutions. To approximately visualise the DNA quantities, all samples were amplified with the 16S rRNA and *nuc*_Hoegh primers in separate PCR reactions and the products were subject to agarose gel electrophoresis.

PCR and agarose gel electrophoresis showed a result for all samples when run with the *nuc*_Hoegh primers (Figure 11), with the BB+P/C samples significantly brighter at the $1x10^{-1}$ and $1x10^{-2}$ dilutions. For PCR with 16S rRNA primers (Figure 12) all bands were fainter, suggesting a less efficient PCR cycle; no product was seen at $1x10^{-2}$, but again the BB+P/C samples were brighter at $1x10^{-1}$.

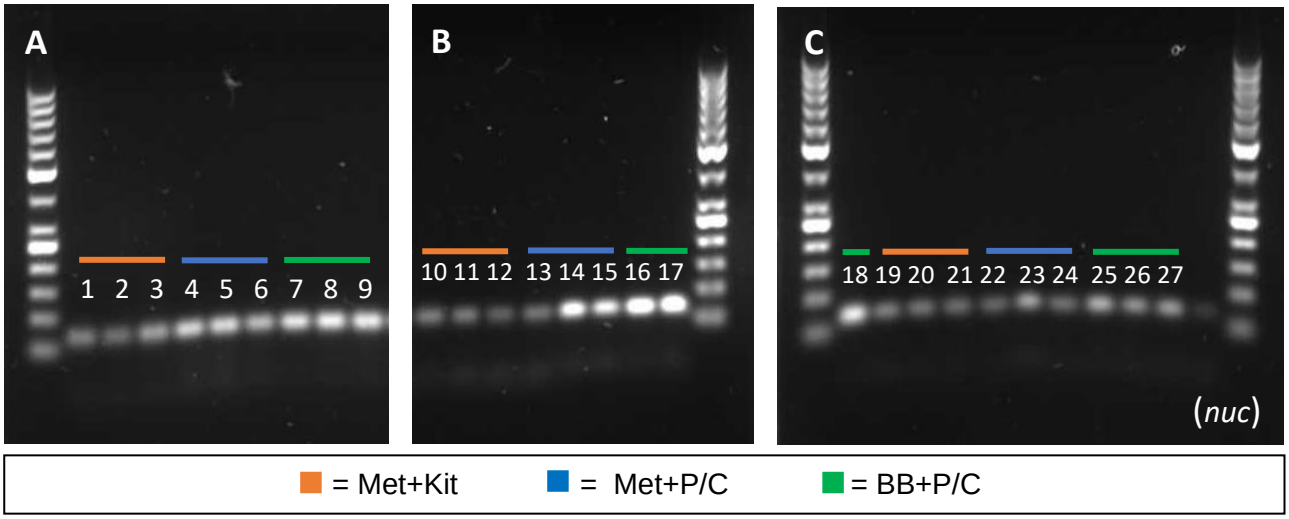


Figure 11 *– Agarose gels (A, B, C) to visualise DNA extracted from serial dilution samples (Neat, 1 in 10, and 1 in 100) and amplified with nuc_Hoegh primers. (Order of bands: Neat samples = Met+Kit (1, 2, 3), Met+P/C (4, 5, 6) and BB+P/C (7, 8, 9). 1 in 10 samples = Met+Kit (10, 11, 12), Met+P/C (13, 14, 15), BB+P/C (16, 17, 18). 1 in 100 samples = Met+Kit (19, 20, 21), Met+P/C = (22, 23, 24), BB+P/C = (25, 26, 27)). Image captured using GeneSys UV transilluminator.*

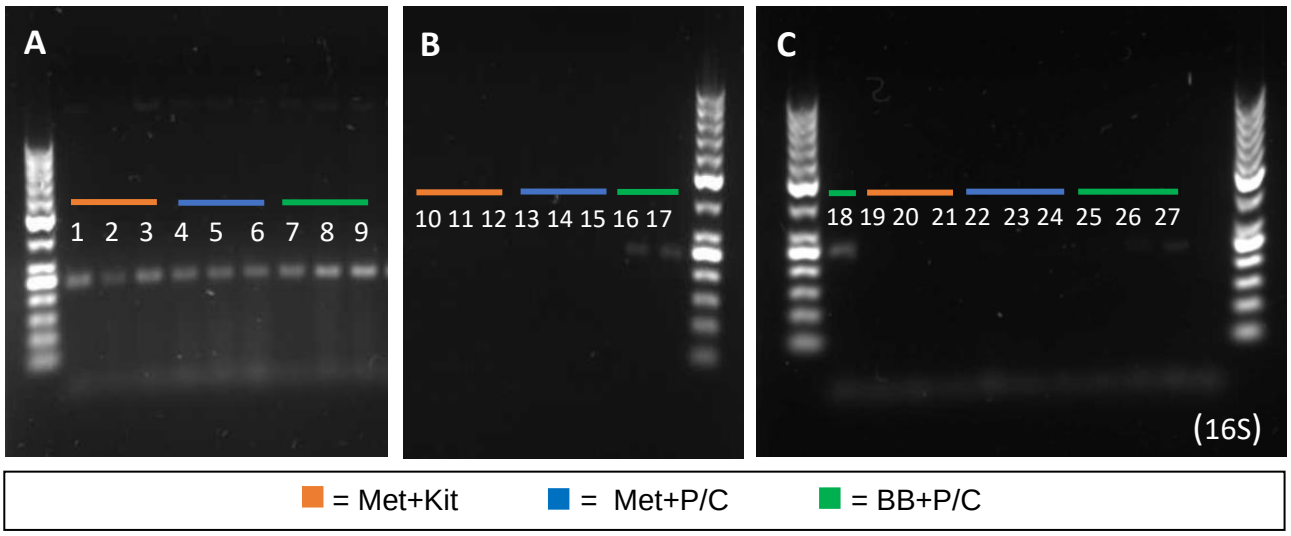


Figure 12 *– Agarose gels (A, B, C) to visualise DNA extracted from serial dilution samples (Neat, 1 in 10, and 1 in 100) and amplified with 16S rRNA primers. (Order of bands: Neat samples = Met+Kit (1, 2, 3), Met+P/C (4, 5, 6) and BB+P/C (7, 8, 9). 1 in 10 samples = Met+Kit (10, 11, 12), Met+P/C (13, 14, 15), BB+P/C (16, 17, 18). 1 in 100 samples = Met+Kit (19, 20, 21), Met+P/C = (22, 23, 24), BB+P/C = (25, 26, 27)). Image captured using GeneSys UV transilluminator.*

As traditional PCR did not have a high enough sensitivity to determine the low quantities of DNA, qPCR was used to assess the relative quantities between samples. DNA from all samples was included in the same qPCR run with 16S rRNA primers. The mean quantitation cycle (Cq) values for each extraction method at each dilution are shown in Table 9 and Figure 13, raw cq values can be found in Appendix 9.9. The p(16S) plasmid standard containing the 16S rRNA gene from *Vibrio natriegens* at a known copy number ($2x10^7$) was included to give an approximate quantification of the amount of DNA in each sample and the expected change in Cq for each dilution.

	Avg. Cq		
	Neat	**1 in 10**	**1 in 100**
Met+Kit	15.67	23.25	28.86
BB+P/C	13.56	18.45	26.13
Met+P/C	19.42	25.59	30.00
p(16S)	13.30	17.40	21.68
Water (NTC)	30.36		

Table 9 *– Data for comparison of methods at neat, 1 in 10, and 1 in 100 dilutions by qPCR using 16S rRNA primers. Data visually represented in Figure 13; raw data in Appendix 9.9.*

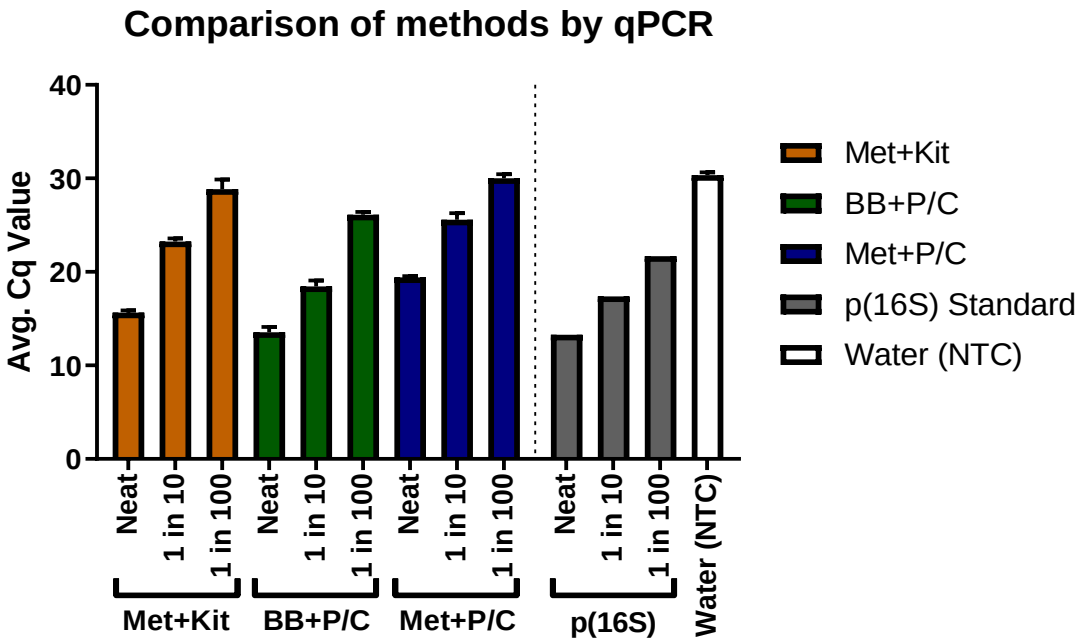


Figure 13 *– Cq values from qPCR using 16S primers on DNA from each extraction method at different dilutions (Neat, 1 in 10, and 1 in 100) against the same dilutions of the p(16S) plasmid standard and a non-template water control. Error bars show standard error. 3 technical repeats were run for each dilution for all extraction methods. Raw Cq values can be found in Appendix 9.9.*

Comparison of the Cq values for each extraction showed that the BB+P/C method was able to extract a higher yield than either the Met+Kit or Met+P/C methods at all dilutions. This led to the selection of the BB+P/C method as the most effective DNA extraction method of the three compared in this study.

5. Results 2 – Development of BB+P/C extraction method

5.1. Phase Lock Gel standardisation experiment

One of the stages in the BB+P/C extraction protocol requires the aqueous phase to be separated from the organic phase below. This stage is reliant on visual determination of when the total aqueous layer has been isolated, creating a potential source of variability due to human error. An error at this stage may lead to loss of DNA through some of the aqueous phase not being isolated and subsequently discarded alongside the organic phase, or may lead to protein contamination of the sample by interrupting the interphase boundary between the aqueous and organic phases. The original protocol for the BB+P/C method which was provided by Dr Michael Cox (University of Birmingham) contained a note recommending the use of Phase Lock Gel to clearly separate the aqueous and organic phases, removing the possibility for human inconsistency to impact the results obtained. Phase lock gel works by producing a barrier layer between the organic and aqueous phases, preventing cross contamination from the organic phase into the aqueous, and removing the human variability factor when isolating the aqueous phase.

To investigate whether phase lock gel would be able to reduce the variability of the BB+P/C DNA extraction method, a liquid culture of *S. aureus* strain JE2 was produced and divided into two sample groups. Each group of samples was subjected to the BB+P/C DNA extraction in parallel, with phase lock gel being used only for the first sample group. Two blank sample sets of TSB were also processed, one with phase lock gel, the other without. This served to show whether the phase lock gel itself would impact the results produced. Three biological repeats were performed, with three technical repeats for each biological repeat. All samples were processed, and the DNA yields of the resulting extracts were quantified using a NanoDrop Spectrophotometer. Results from this phase lock gel experiment are shown in Table 10 and Figure 14, raw data can be found in Appendix 9.10.

Sample		Avg. DNA quantity (ng/µl)	Standard deviation
JE2	PLG -	1425.01	396.15
	PLG +	2014.81	115.81
Blank (TSB)	PLG -	19.54	4.12
	PLG +	18.82	4.57

***Table 10** – Average DNA yields from phase log gel (PLG) tests to determine the effect of PLG on the variability of samples. DNA was quantified using a NanoDrop spectrophotometer. Data visually represented in Figure 14. Raw data can be found in Appendix 9.10.*

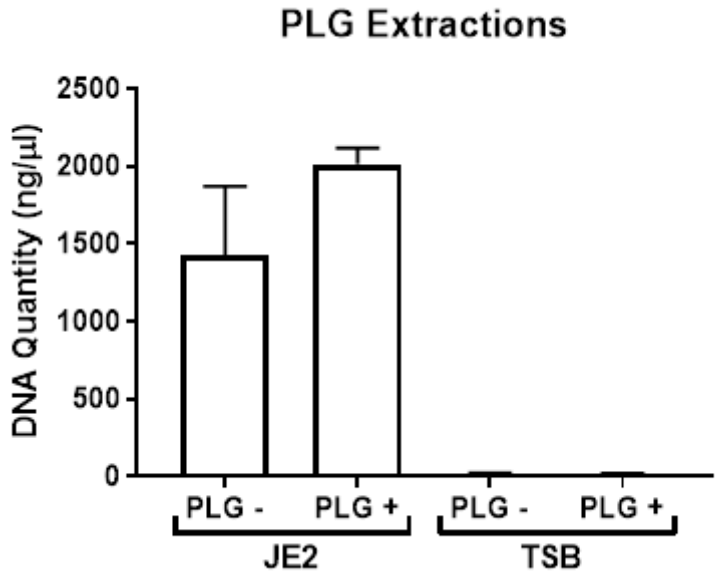


***Figure 14** – Average DNA yields from phase log gel (PLG) tests to determine the effect of PLG on the variability of samples. For each condition, 3 samples were measured in triplicate using a NanoDrop spectrophotometer. Error bars show standard deviation. Raw data can be found in Appendix 9.10.*

For each sample in this experiment, three readings were taken on a NanoDrop spectrophotometer and an average reading was produced for each sample (1 to 12), as well as an average quantity for each sample type (JE2, JE2+PLG, TSB and TSB+JE2). The standard deviation was calculated from the raw data readings to show the total variation in each sample type. This experiment showed that when samples were processed using phase lock gel, the variability of the samples was significantly reduced. The JE2 samples processed without phase lock gel showed a standard deviation of 396.15ng/µl between samples, compared to 115.81ng/µl in the JE2+PLG samples; almost a 3.5-fold difference between sample types.

Furthermore, both the TSB blank samples showed comparable results with no significant difference between the TSB and TSB+PLG sample types (P=0.729 in a 2-tailed Students t-test), demonstrating that the presence of phase lock gel does not contaminate the sample and artificially increase the yield detected. Interestingly, the presence of phase lock gel caused a significant increase in the amount of DNA obtained from the samples of JE2 (P=0.00186 in a 2-tailed Students t-test with unequal variance). This increase was not observed in the TSB+PLG blank samples when compared to the TSB samples not containing PLG. Therefore, it can be assumed that the increase in DNA obtained is not due to contamination from the phase lock gel itself, or from an interference from residual phase lock gel on the A260/280 reading of the spectrophotometer used to quantify DNA yield. Aside from the *S. aureus* JE2 culture used, there was no other source of DNA in the experiment, suggesting that the yields produced by the JE2+PLG samples more closely reflect the true yield of the sample and that there is a significant loss of yield (on average) in the samples processed without using phase lock gel. Closer scrutiny of the samples supports this theory as sample 1 (JE2 without PLG) is comparable to the JE2+PLG samples, but the other two replicate JE2 samples processed without phase lock gel show a reduced DNA yield. This may be due to either not isolating the entirety of the aqueous phase, thereby losing some of the DNA in the samples during processing, or it may be due to contamination of the sample from material in either the interphase layer or organic phase which could interfere with the A260/280 reading from the NanoDrop spectrophotometer used to quantify each sample.

5.2. Production of *nuc*_qPCR primers and p(*nuc*) standard

Prior to performing the extended serial dilution experiments to determine the limit of detection (LoD) of the BB+P/C DNA extraction method, it was necessary to make the protocol more specific for detecting *S. aureus* from complex samples. This included producing a set of primers specific to the *nuc* gene which wouldn't produce a product with DNA templates from any other microorganisms (*nuc*_qPCR primers), and also included the production of a plasmid standard (p(*nuc*)) in order to determine the LoD of the qPCR analysis stage with the *nuc* primers.

New primers for the *nuc* gene were required for extraction from complex samples as the *nuc*_Hoegh set had been shown to produce a product with *Streptococcus pneumoniae* strain d39 in an experiment unrelated to this project. Until this point of the project, the *nuc*_Hoegh primers fulfilled the requirements necessary of the *nuc* primer set as all experiments had been performed on pure cultures of *S. aureus*, however the *nuc*_Hoegh primers would not be suitable for use on complex samples of unknown microbial compositions.

The first stage in producing the new *nuc*_qPCR primers was to perform multiple sequence alignment from a broad selection of *S. aureus* strains to determine a consensus sequence for the *nuc* gene. Sequences for the *nuc* gene were taken for 31 strains of *S.* aureus (Table 11); the gene was inverted in several strains so the reverse complement for these strains was produced using an online sequence converter [50]. Multiple sequence alignment was then performed using MultAlin [51] producing a consensus sequence for the *S. aureus nuc* gene (Figure 15). The most homologous region of the *nuc* gene consensus sequence was then selected as a target sequence for the *nuc*_qPCR primers (Table 12).

NCBI accession number	***S. aureus* Strain**
NC_007793	USA300_FPR3757
NC_010079	USA300_TCH1516
NC_002951	COL
NC_007795	NCTC_8325
NC_016912	VC40
NC_009641	Newman
NC_017331	TW20
NC_017341	JKD6008
NC_017347	T0131
NC_017351	11819_97
NC_002953	MSSA476
NC_003923	MW2
NC_013450	ED98
NC_017343	ECT_R_2
NC_009487	JH9
NC_009632	JH1
NC_017340	04_02981
NC_002745	n315
NC_002758	Mu50
NC_009782	Mu3
NC_016928	M013
NC_017349	LGA251
NC_007622	RF122
NC_017337	ED133
NC_017338	JKD6159
NC_017763	HO_5096_0412

NCBI accession number	*S. aureus* Strain
NC_002952	MRSA252
NC_017342	TCH60
NC_017673	71193
NC_017333	ST398
NC_018608	08BA02176

Table 11 *– List of 31 representative strains of S. aureus used for multiple sequence alignment of nuc gene as part of the production of nuc_qPCR primers. List originally produced by AureoWiki [52] to produce a pan-genome for S. aureus.*

***nuc* gene consensus sequence:**

ATGACAGAATACTTATTAAGTGCTGGCATATGTATGGCAATTGTTTCAAT
ATTACTTATAGGGATGGCTATCAGTAATGTTTCGAAAGGGCAATACGCAA
AGAGGTTTTTCTTTTTCGCTACTAGTTGCTTAGTGTTAACTTTAGTTGTA
GTTTCAAGTCTAAGTAGCTCAGCAAATGCATCACAAACAGATAACGGCGT
AAATAGAAGTGGTTCTGAAGATCCAACAGTATATAGTGCAACTTCAACTA
AAAAATTACATAAAGAACCTGCGACATTAATTAAAGCGATTGATGGTGAT
ACGGTTAAATTAATGTACAAAGGTCAACCAATGACATTCAGACTATTATTG
GTTGATACACCTGAAACAAAGCATCCTAAAAAAGGTGTAGAGAAATATGG
TCCTGAAGCAAGTGCATTTACGAAAAAAATGGTAGAAAATGCAAAGAAAA
TTGAAGTCGAGTTTGACAAAGGTCAAAGAACTGATAAATATGGACGTGGC
TTAGCGTATATTTATGCTGATGGAAAAATGGTAAACGAAGCTTTAGTTCG
TCAAGGCTTGGCTAAAGTTGCTTATGTTTATAAACCTAACAATACACATG
AACAACTTTTAAGAAAAAGTGAAGCACAAGCGAAAAAAGAGAAATTAAAT
ATTTGGAGCGAAGACAACGCTGATTCAGGTCAATAA

Figure 15 *– Full consensus sequence of nuc gene produced by multiple sequence alignment of 31 representative strains of S. aureus. Red text denotes common SNPs.*

Most homologous region of *S. aureus nuc* gene:		AAAATTGAAGTCGAGTTTGACAAAGGTCAAAGAACTGATAAATATG GACGTGGCTTAGCGTATATTTATGCTGATGGAAAAATGGTAAACGA AGCTTTAGTTCGTCAAGGCTTGGCTAAAGTTGCTTATGTTTATAAAC CTAACAATACACATGAACAACTTTTAAGAAAAAGTGAAGCACAAGC
Primers selected:	*nuc*_qPCR_F:	ATTGAAGTCGAGTTTGACAAAG
	*nuc*_qPCR_R:	TTGTGCTTCACTTTTTCTTAAAAG

Table 12 *- The most homologous region of the nuc gene, determined by multiple sequence alignment of the nuc gene from 31 representative strains of S. aureus. Primers were produced to amplify within this homologous region in order to attain the highest level of coverage of S. aureus strains. The target sequence for each primer is shown in red.*

The specificity of the primers produced (*nuc*_qPCR_F and *nuc*_qPCR_R) was assessed using the NCBI Primer BLAST tool [53], this found the primers to only have an expected specificity to the *S. aureus nuc* gene. As the previous *nuc* primer set (*nuc*_Hoegh) had produced a product with *S. pneumoniae* strain d39, a specificity test was performed including 2 genomic DNA (gDNA) templates from *S. aureus* strain JE2, 2 from *S. pnumoniae* strain d39, and 2 non-template controls of qPCR-grade water. All samples were analysed by qPCR using the *nuc*_qPCR primers. A product was produced with the JE2 template, but not with the d39 or water samples (Figure 16 and Table 13).

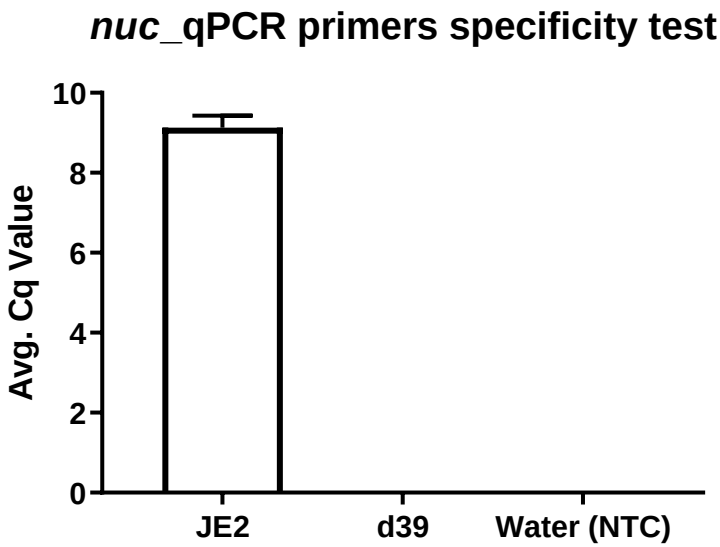


Figure 16 *– Results from nuc_qPCR primers specificity test. qPCR performed on samples containing template DNA from S. aureus JE2, S. pneumoniae d39 or water non-template control using nuc_qPCR primers. Error bars show standard error. For each condition, N = 2. Data shown in Table 13.*

Sample	Cq Value		
	A	B	Avg.
JE2	9.34067	8.92179	9.13
d39	-1	-1	0.00
NTC (water)	-1	-1	0.00

Table 13 *– Data from nuc_qPCR primers specificity test. qPCR performed on samples containing template DNA from S. aureus JE2, S. pneumoniae d39 or water non-template control using nuc_qPCR primers. Demonstrated specificity of primers for S. aureus.*

To determine the efficiency of the *nuc*_qPCR primers, a 1 in 5 serial dilution was performed on JE2 gDNA from 62.5ng/µl to 0.5ng/µl; 2 technical replicates per dilution were used for qPCR as well as 2 non-template control samples (Table 14a). From the Cq values acquired, it was possible to calculate the efficiency of the primer set as 94.3% with an R2 value of 0.9994 (Table 14b).

Sample (JE2 DNA)	Cq Value		
	A	B	Avg.
62.5ng	11.16032	10.77660	10.97
12.5ng	13.51483	13.44775	13.48
2.5ng	17.32561	14.60272	15.96
0.5ng	17.31467	19.12055	18.22
NTC (water)	-1	-1	0.00

Analysis	
Slope	-3.466575976
R^2	0.9994
Efficiency (%)	94.30

***Table 14** – Efficiency of nuc_qPCR primers. A) Cq values from qPCR on a serial dilution series of S. aureus JE2 genomic DNA and water non-template control. B) Determined efficiency of nuc_qPCR primer set based on Cq values obtained. R^2 value indicates accuracy of efficiency calculation (optimal value: 1); Slope is used to determine efficiency value (optimal value: -3.3).*

A new plasmid standard was required for use with the *nuc* primer set as the LoD for the method is likely to differ depending on the primer set used, and therefore each primer set must have its own complimentary plasmid standard. The p(*nuc*) standard consists of the full length *nuc* gene from *S. aureus* in a pCDF-Duet-1 vector, diluted to a known copy number of $2x10^7$ copies per µl. It was produced by isolating the *nuc* gene from *S. aureus* strain JE2 and inserting it into the plasmid by restriction digestion. The p(*nuc*) plasmid was then transformed into DH5α *E. coli* cells, checked by colony PCR and cultured in liquid broth.

The p(*nuc*) plasmid was then extracted using a plasmid miniprep kit and diluted to $2x10^7$ copies per µl. To monitor the success of the cloning procedure, the DNA products were inspected by agarose gel electrophoresis at several checkpoints. These checks were performed following isolation of the *nuc* gene (Figure 17), following the restriction digest of the *nuc* fragment into the pCDF-Duet-1 plasmid vector (Figure 18), and following the transformation of the p(*nuc*) plasmid into DH5α (Figure 19).

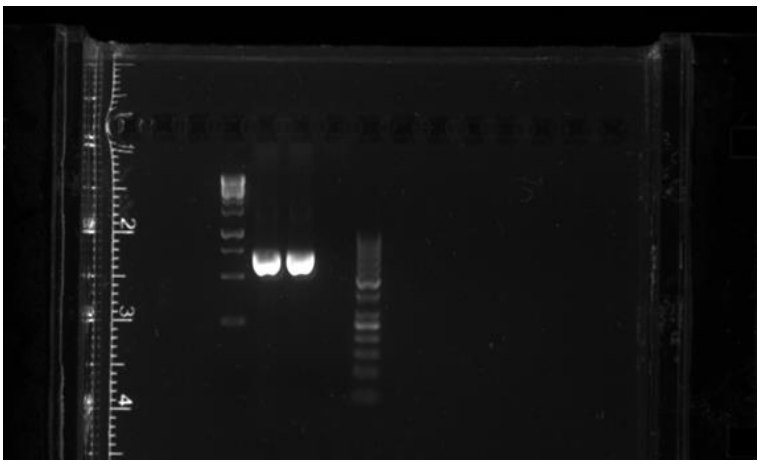


Figure 17 *– Agarose gel to check nuc fragment amplified by PCR. Electrophoresis conditions: 300mA, 90V, 30m. 1.5% agarose gel containing 0.1µl/ml SYBR-Safe gel stain. Total volume of 5µl loaded per well. Order of wells: 1kb ladder, nuc fragment template positive 1 (nuc +ve 1), nuc fragment template positive 2 (nuc +ve 2), nuc fragment template negative (nuc -ve), 50bp ladder. Gel confirmed the presence of a fragment of the expected size in the template positive samples (nuc +ve 1 and nuc +ve 2), which was absent in the negative sample (nuc -ve).*

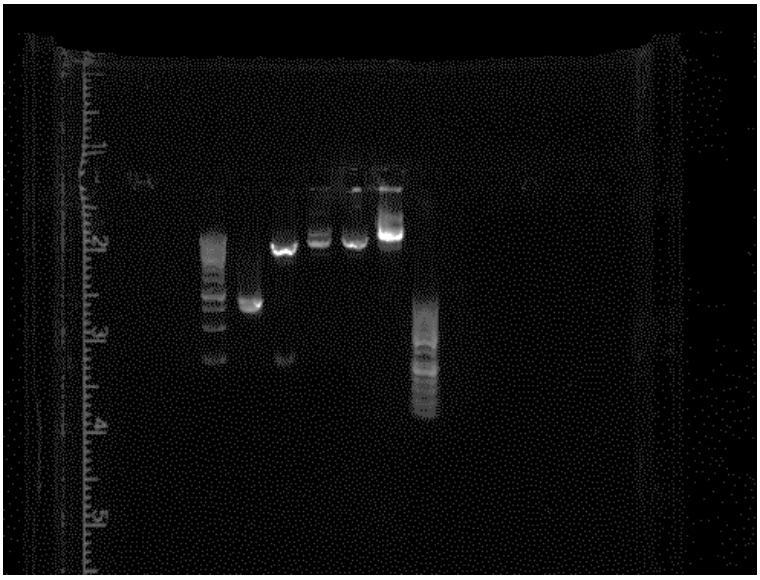


Figure 18 *– Agarose gel to check restriction digest of pCDF-Duet-1 plasmid vector and nuc fragment using EcoRI and KpnI. Electrophoresis conditions: 300mA, 90V, 60m. 1% agarose gel containing 0.04µl/ml Gel-Red stain. Total volume of 5µl loaded per well. Order of wells: 1kb ladder, nuc gene double digest (EcoRI + KpnI), pCDF-Duet-1 double digest (EcoRI + KpnI), pCDF-Duet-1 single digest (EcoRI), pCDF-Duet-1 single digest (KpnI), pCDF-Duet-1 uncut plasmid, 50bp ladder. This confirmed that the restriction digest was successful, the nuc double digest showed a band at the expected size, and all plasmid digests showed bands as expected relative to each other and the uncut plasmid.*

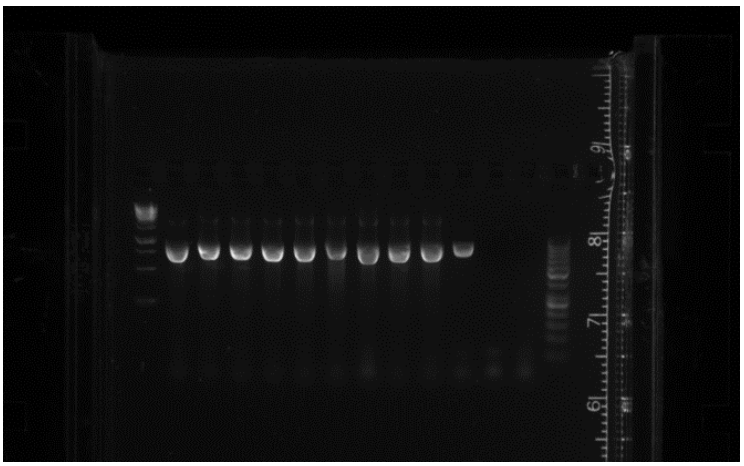


Figure 19 *– Agarose gel of colony PCR product to check transformation of p(nuc) (pCDF-Duet-1 plasmid vector and nuc fragment) into DH5α cells. Electrophoresis conditions: 300mA, 90V, 60m. 1% agarose gel containing 0.1µl/ml SYBR-Safe gel stain. Total volume of 7.5µl loaded per well. Order of wells: 1kb ladder, p(nuc) positive DH5α 1 to 10, p(nuc) negative DH5α 1 to 2, 50bp ladder. This confirmed that all 10 positive colonies of DH5α did contain p(nuc), and that both negative colonies of DH5α did not contain p(nuc). The negative colonies produced no bands as the forward primer used for PCR (ACYCDuetUP1) was specific to the plasmid and the reverse primer (RD_nuc_R_(KpnI)) was specific to the nuc insert. Therefore, only colonies containing the nuc gene in the plasmid would produce a result.*

5.3. Extended serial dilution experiments to determine limit of detection (LoD)

5.3.1. LoD of qPCR analysis stage

The serial dilution comparison previously carried out (Section 4.4; neat to $1x10^{-2}$) was enough to see a difference between the methods and determine which one had the highest sensitivity of the three. However, in order to assess the extent of the most effective method's sensitivity, a further serial dilution experiment including samples ranging from a neat culture to 1 CFU/ml was required. It would not have been appropriate to perform this extended serial dilution with all three methods in parallel as this would have been highly laborious, decreasing the quality of the results obtained.

As the BB+P/C method had been shown to be the most effective method at extracting microbial DNA from low yield samples, it was necessary to find the limit of detection (LoD) of this method with both the *nuc* primers and the 16S rRNA primers. The BB+P/C method consists of two stages which may affect LoD: the DNA extraction stage, and the qPCR analysis stage. As qPCR is a notoriously sensitive molecular detection technique it was expected that if the LoD was impaired by

this method, most of the impact would be from the DNA extraction stage itself. As investigation of the DNA extraction stage is reliant on qPCR analysis, the qPCR stage was investigated first in isolation.

To find the baseline LoD of the qPCR analysis stage, the LoD for the p(*nuc*) and p(16S) standards was determined through analysis in independent qPCR reactions. Each standard at $2x10^7$ copies/µl was first diluted to $1x10^7$ copies/µl, then a 7-fold, 1 in 10 serial dilution was performed to produce samples containing a range from 10,000,000 copies per µl to 1 copy per µl. From each of these samples, two samples of 2µl each were then analysed by qPCR. In each qPCR run, two 2µl water non-template controls (NTC's) were also run alongside the standards; these were identical to the other samples except they contained 2µl of qPCR grade water rather than a DNA template. Average Cq values are shown in Table 15, raw data can be found in Appendix 9.11.1.

Dilution of Plasmid Standard	Copy Number used	Avg. Cq value	
		p(*nuc*)	p(16S)
N	**2×10^7**	14.25	13.73
1×10^{-1}	**2×10^6**	17.4	17.54
1×10^{-2}	**2×10^5**	21.18	21.05
1×10^{-3}	**2×10^4**	23.99	23.97
1×10^{-4}	**2×10^3**	26.87	27.61
1×10^{-5}	**200**	29.07	29.83
1×10^{-6}	**20**	31.19	31.06
1×10^{-7}	**2**	0	31.86
Water NTC	**0**	0	31.93

***Table 15** – Average Cq values of serial dilution samples of plasmid standards, p(nuc) and p(16S), analysed by qPCR to find the LoD of the qPCR analysis stage in isolation. 2µl of each sample at $1x10^{(n)}$ was used per qPCR reaction. Visually represented in Figure 20. Raw data in Appendix 9.11.1.*

Through comparison of the Cq values produced for each serially diluted standard, it was possible to calculate the LoD of the qPCR analysis stage for each primer set. The cut-off point used to determine a significant difference between a sample and the water NTC was 1Cq + 2 standard deviations (1Cq+2SD). The *nuc* primers did not produce a detectable product in the water NTC or the sample containing 2 copies of p(*nuc*) in the qPCR run, so the LoD when using the *nuc*_qPCR primers and 40 PCR cycles was determined as the lowest dilution to produce a result. This was 20

copies of the *nuc* gene (Figure 20a). The LoD ascertained for qPCR analysis with the 16S rRNA primers was determined by taking the first dilution to not be within 1Cq+2SD of the water NTC, this was 200 copies of the 16S rRNA gene (Figure 20b). Therefore, the qPCR analysis stage was determined to have a limit of detection of 20 copies with the *Staphylococcus aureus* specific *nuc* primers and 200 copies with the universal 16S rRNA primers; these LoD scores for the qPCR stage alone were used as a baseline for evaluating the LoD of the DNA extraction and qPCR analysis stages combined.

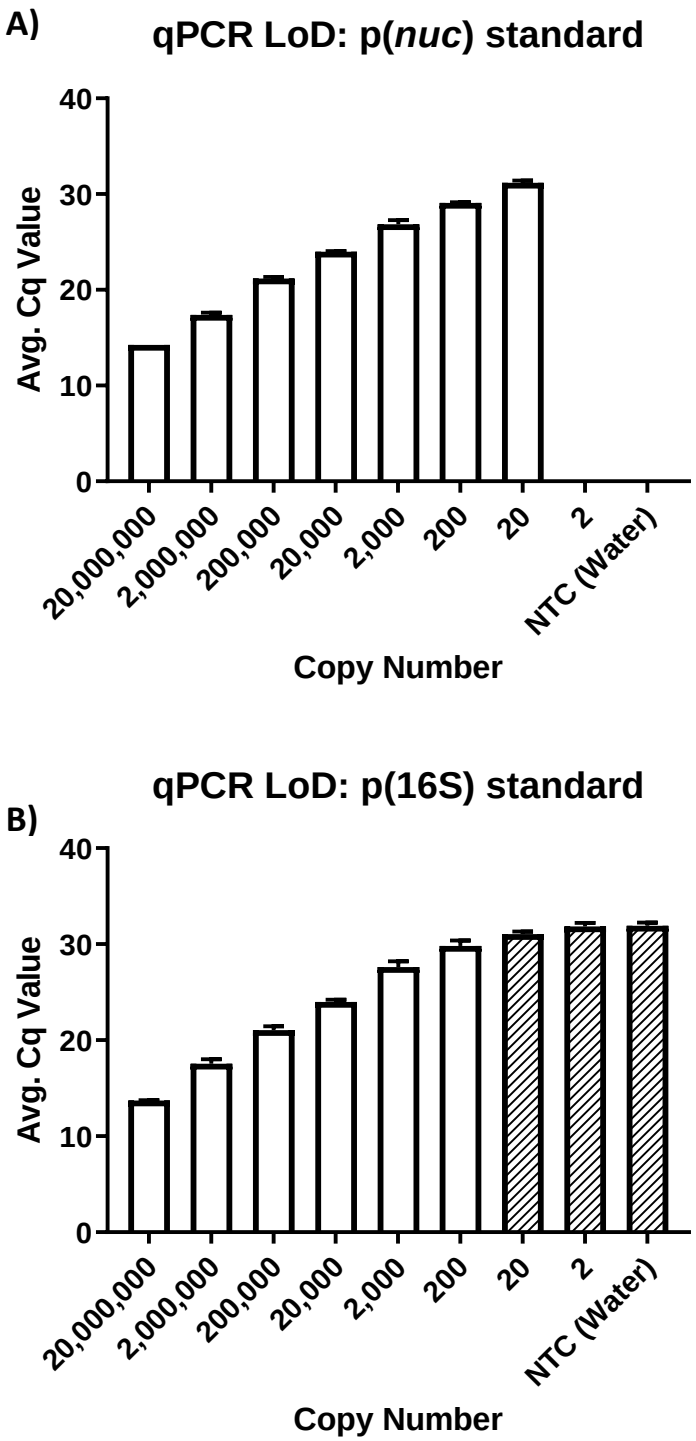


Figure 20 *– Graphs to show avg Cq values from qPCR analysis of a) p(nuc) and b) p(16S) standard serial dilutions to calculate LoD of qPCR stage. Hashed lines denote samples not significantly different to the water non-template control (i.e. avg. Cq of sample is within 1Cq+2SD of NTC). Limit of detection determined as the last sample to produce an average Cq statistically distinct from the average Cq of the NTC. A) avg. qPCR values for serially diluted p(nuc) standard; B) avg. qPCR values for serially diluted p(16S) standard. Error bars show standard error. Summary data can be found in Table 15 and raw data can be found in Appendix 9.11.1.*

5.3.2. LoD of BB+P/C DNA extraction and qPCR analysis stages combined

To determine the LoD of the combined extraction and qPCR analysis stages, an extended 1 in 10 serial dilution was carried out on an overnight culture of *S. aureus* JE2. The dilution series was performed from a neat culture down to $1x10^{-14}$ to ensure it would extend past a cell count of 1 CFU/ml. DNA was extracted from each of these samples in triplicate, alongside three blank samples of TSB using the BB+P/C method. Following the extractions, samples were compared by qPCR analysis in order to determine the LoD for each primer set under the conditions used: each sample had a starting volume of 1ml, DNA was resuspended in 60µl after extraction, and 2µl of each DNA extract was used for qPCR analysis.

By the $1x10^{-9}$ dilution of the serial dilution series, the sample contained approximately 1 CFU/ml (Table 16, raw data in Appendix 9.11.2). Scaling this up across the dilution series indicates the neat starting culture of JE2 contained approximately $1x10^{9}$ CFU/ml. In a clinical setting, a positive BAL sample is determined as one which contains more than $1x10^{4}$ CFU/ml [17], this would equate to the $1x10^{-5}$ dilution in this experiment.

Sample		Avg. CFU/ml	Estimated CFU/ml
Number	**Dilution**		
1	Neat	-	~1,000,000,000 (1×10^{9})
2	1×10^{-1}	-	~100,000,000 (1×10^{8})
3	1×10^{-2}	-	~10,000,000 (1×10^{7})
4	1×10^{-3}	-	~1,000,000 (1×10^{6})
5	1×10^{-4}	-	~100,000 (1×10^{5})
6	1×10^{-5}	-	~10,000 (1×10^{4})
7	1×10^{-6}	1032.00	~1,000 (1×10^{3})
8	1×10^{-7}	119.00	~100 (1×10^{2})
9	1×10^{-8}	10.00	~10
10	1×10^{-9}	0.33	~1
B	Blank (TSB)	0.00	0

***Table 16** – Average CFU/ml plate counts from a serial dilution of JE2 liquid culture used for LoD experiments. Approximate CFU/ml counts for later dilutions ($1x10^{-6}$ to $1x10^{-8}$) were extrapolated to predict the estimated CFU/ml of earlier dilutions in the series. Raw data can be found in Appendix 9.11.2.*

For each dilution containing at least 1 CFU/ml (samples 1 to 10; neat culture to the $1x10^{-9}$ dilution), total DNA was extracted from three individual 1ml samples using the BB+P/C method. Three blank samples of TSB were also put through the extraction procedure parallel to the serial dilution samples. After the extraction, DNA was resuspended in 60µl of low-EDTA TE buffer. 2µl from each 60µl DNA extract was analysed via qPCR using both generic 16S rRNA primers, and *Staphylococcus aureus* specific *nuc* primers. Comparison of the Cq values produced by serially diluted samples against the blank TSB sample would show the limit of detection (LoD) of the extraction method and qPCR analysis using both sets of primers. Ideally, the standards would have also been run alongside the DNA extracts from samples containing JE2, however this was not possible due to space constraints on the thermocycler.

The qPCR analysis of the DNA extracts produced a series of Cq values which were used to calculate the LoD of the method for each primer set. This was achieved by selecting the last dilution before the point at which the average Cq value from the DNA extract would be within 1 Cq and 2 standard deviations (1Cq+2SD) from either the blank sample or non-template control (NTC). The LoD of the BB+P/C DNA extraction method and qPCR analysis combined was found to be $1x10^{3}$ CFU/ml with *nuc* primers (Figure 21a), and $1x10^{4}$ CFU/ml with 16S rRNA primers (Figure 21b) under the conditions stated. These LoD cut-off points indicate the lowest number of cells in 1ml of a sample which can be extracted using the BB+P/C method and reliably detected through qPCR analysis, producing a result which is clearly distinct from an unspiked control sample with each primer set. Average Cq values are shown in Table 17 and raw Cq values can be found in Appendix 9.11.3. A summary of all LoD's calculated can be found in Table 18.

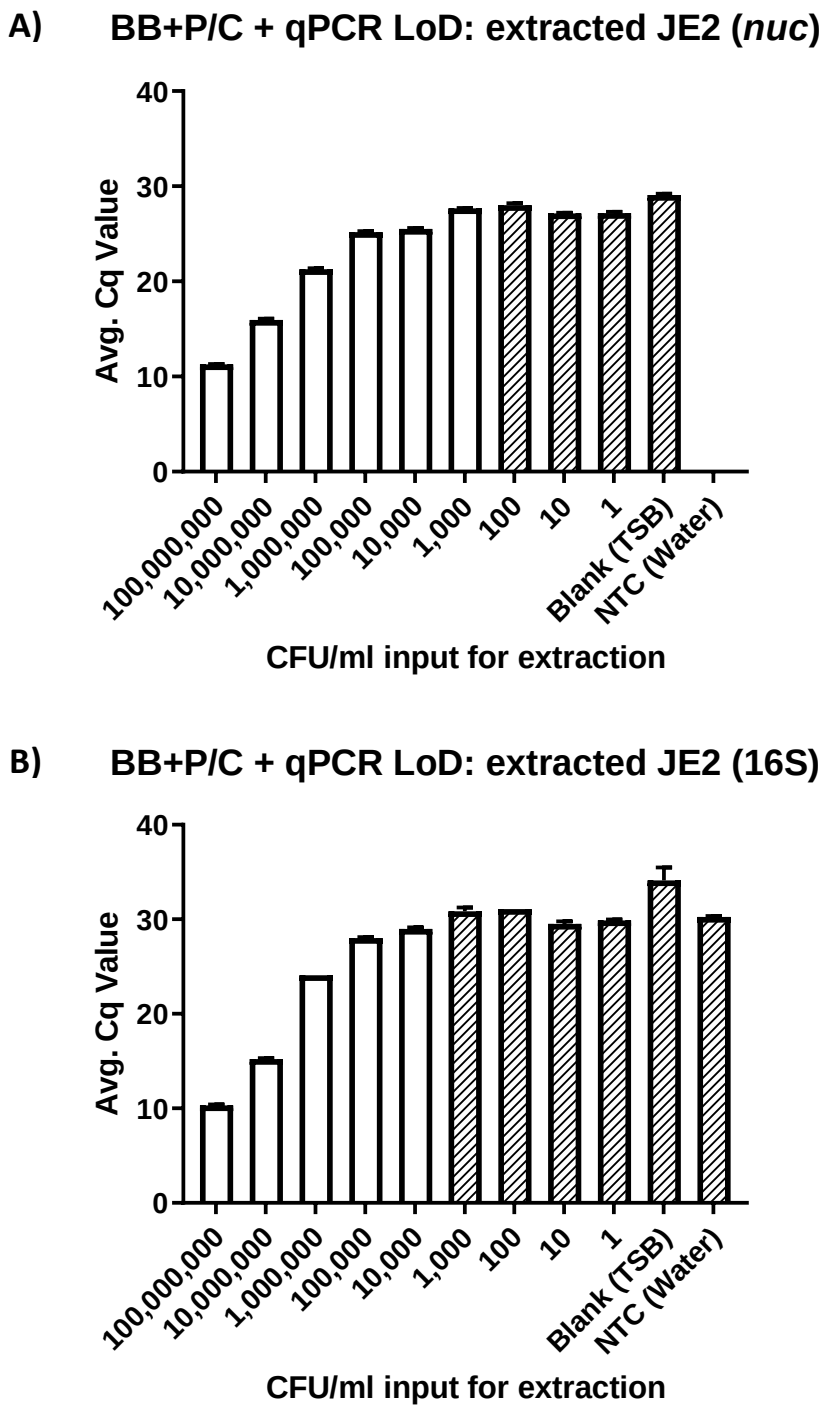


Figure 21 *– graphs to show avg Cq values used to calculate LoD of BB+P/C DNA extraction + qPCR analysis. DNA extracts used for qPCR were from a serial dilution series on S. aureus JE2. Hashed lines denote samples not significantly different to the blank (TSB) control (i.e. avg. Cq of sample is within 1Cq+2SD of blank (TSB) control). Limit of detection determined as the last sample to produce an average Cq statistically distinct from the average Cq of the blank (TSB). A) avg. qPCR values using nuc_qPCR primers; B) avg. qPCR values using 16S rRNA primers. Error bars show standard error. Summary data can be found in Table 17 and raw data can be found in Appendix 9.11.3.*

		Avg. Cq value	
		nuc	16S
CFU/ml of JE2	100,000,000	11.30	10.32
	10,000,000	15.95	15.21
	1,000,000	21.29	24.09
	100,000	25.19	28.02
	10,000	25.53	29.00
	1,000	27.69	30.85
	100	28.02	31.07
	10	27.16	29.51
	1	27.19	29.89
Blank (TSB)		29.10	34.11
NTC (Water)		0.00	30.26

Table 17 – Average Cq values from qPCR to determine the LoD of BB+P/C and qPCR stages combined. qPCR performed using both 16S rRNA primers and nuc primers. Visually represented in Figure 21. Raw data in Appendix 9.11.3.

		Limit of detection (LoD)
qPCR analysis only	**p(*nuc*) standard**	20 copies
	p(16S) standard	200 copies
DNA extraction + qPCR analysis	**JE2 (*nuc* primers)**	1 x 10^3 CFU/ml
	JE2 (16S rRNA primers)	1 x 10^4 CFU/ml

***Table 18** – LoD calculated for each primer set (nuc_qPCR and 16S rRNA) for either qPCR analysis stage only, or for DNA extraction and qPCR analysis stages combined.*

It is important to note that the LoD figures for the extraction and qPCR analysis combined are based on Cq results from qPCR runs utilising only 2µl of the total 60µl DNA extract for each sample. The 60µl extract theoretically contains the total quantity of DNA present in the original 1ml sample. Loading the full 60µl into the qPCR reaction would overload the qPCR run preventing the acquisition of accurate results; therefore, 2µl was used in each qPCR run and the Cq values produced are representative of only 1/30th of the initial 1ml sample. For this project, an LoD has been calculated based on the input number of cells for each sample to show the lowest number of cells this method is able detect from 1ml of a sample when DNA is resuspended in 60µl and 2µl of this is then used for qPCR analysis. Further considerations regarding the quantity of DNA represented in the qPCR analysis stage are explored in more detail in the discussion section.

5.4. DNA extractions from spiked clinical samples of pleural fluid

5.4.1. Pleural fluid samples

Once the capabilities of the BB+P/C DNA extraction protocol had been established on mock-samples of a single organism and the LoD was found to be acceptable for the intended uses of this method, it was important to ensure the efficacy would not be impaired when used on complex clinical samples. For these experiments, 30ml of pleural fluid was obtained from a lymphoma patient at Southmead Hospital in Bristol. This patient suffered pleural effusion requiring pleurocentesis, the fluid collected was frozen and transported to the University of Bristol where it was stored at -20°C until required.

The pleural fluid was defrosted on ice for 45m, then 30ml was isolated and vortexed to homogenise any clumps of cells before being separated into six individual 5ml samples. A liquid culture of *S. aureus* strain JE2 was grown overnight and used as the neat culture for a serial dilution series; the neat culture was shown to be at $1.83x10^9$ CFU/ml by culturing spread plates of each dilution and extrapolating from the discernible values at lower dilutions (Table 19).

Dilution	CFU/ml			Avg. CFU/ml
	plate 1	plate 2	plate 3	
Neat	TNTC	TNTC	TNTC	**~1.83 x 10^9**
1 x 10^{-1}	TNTC	TNTC	TNTC	**~1.83 x 10^8**
1 x 10^{-2}	TNTC	TNTC	TNTC	**~1.83 x 10^7**
1 x 10^{-3}	TNTC	TNTC	TNTC	**~1.83 x 10^6**
1 x 10^{-4}	TNTC	TNTC	TNTC	**~1.83 x 10^5**
1 x 10^{-5}	TNTC	TNTC	TNTC	**~1.83 x 10^4**
1 x 10^{-6}	1752	1792	1880	**1808.00**
1 x 10^{-7}	168	212	203	**194.33**
1 x 10^{-8}	22	16	14	**17.33**
1 x 10^{-9}	0	3	1	**1.33**
1 x 10^{-10}	0	0	0	**0.00**

Table 19 *– Plate count values of colony forming units per ml (CFU/ml) from serial dilution series of S. aureus JE2 liquid culture. Average CFU/ml for samples from $1x10^5$ to the neat sample are predicted values which were extrapolated from the values counted at lower levels of the dilution series ($1x10^{-6}$ to $1x10^{-8}$).*

Half of the six 5ml samples of pleural fluid were then spiked with 10µl of the neat JE2 liquid culture, an equivalent of $1.83x10^7$ cells per sample; 10µl of sterile TSB was added to the remaining pleural fluid samples. Alongside the pleural fluid samples, six control samples of phosphate buffered saline (PBS) were also used, half of which were also spiked with $1.83x10^7$ cells of JE2 and half with 10µl of sterile TSB. The 4 sample types produced are summarised in Table 20. All samples were put through the BB+P/C double extraction procedure, producing 60µl of DNA extract for each sample which could be analysed by qPCR using two primer sets: one targeting the *Staphylococcus aureus* specific *nuc* gene, and one targeting the V3-V4 region of the 16S rRNA gene which is universal to all bacteria. There are two stages during the sample processing where the quality of results could be impaired, these are either during the extraction itself, or during the analysis by qPCR. As the extraction stage is reliant on qPCR for analysis of results, it was logical to investigate the impacts of using a complex clinical sample on the qPCR stage in isolation before investigating the efficacy of the two stages combined.

	Sample			
	1 (PBS+TSB)	2 (PBS+JE2)	3 (PF+TSB)	4 (PF+JE2)
Pleural fluid (+) or PBS (-)	-	-	+	+
10µl *of S. aureus* (JE2) (+) or 10µl TSB (-)	-	+	-	+

Table 20 *– Summary of sample types produced for pleural fluid spiking experiments. All samples were used for JE2 spike experiments to determine the efficacy of the BB+P/C DNA extraction method when extracting from a complex clinical sample. Sample 3 was also used for qPCR inhibitor checks analysing the impact of PF on the qPCR analysis stage. The 10µl of JE2 added to samples 2 and 4 contained ~$1.83x10^7$ cells.*

5.4.2. qPCR inhibition test: *nuc_qPCR* primers

The robustness of the qPCR stage when using a template derived from a clinical sample was examined through a set of spike-in experiments for each primer set. The clinical samples of pleural fluid were likely to contain a background level of microorganisms which would produce a product with the 16S rRNA primers, so the *Staphylococcus aureus* specific *nuc* primers were tested first.

Four samples were produced of the p(*nuc*) standard at $2x10^7$ copies per µl; these were then separated into two individual sets. Each set also contained two non-template control samples of qPCR grade water. To the first set, 1µl of DNA extract from the un-spiked pleural fluid sample (PF+TSB) was added to each sample. To the second set, 1µl of qPCR grade water was added. qPCR was then performed on two technical repeats for each sample (Figure 22; raw data in Appendix 9.12.1).

If present, any PCR inhibitors in the pleural fluid DNA extract would impact the acquisition of accurate and reliable results through qPCR analysis on the samples containing DNA derived from the clinical pleural fluid samples. Therefore, if the samples containing pleural fluid produce an equivalent Cq value to the samples without pleural fluid, it can be assumed that any inhibitors present do not have a significant impact on the results from this qPCR run. However, this does not prove a definitive absence of any inhibitors and this must be taken into consideration if DNA extracts in future studies are used for any further downstream functions, such as sequencing.

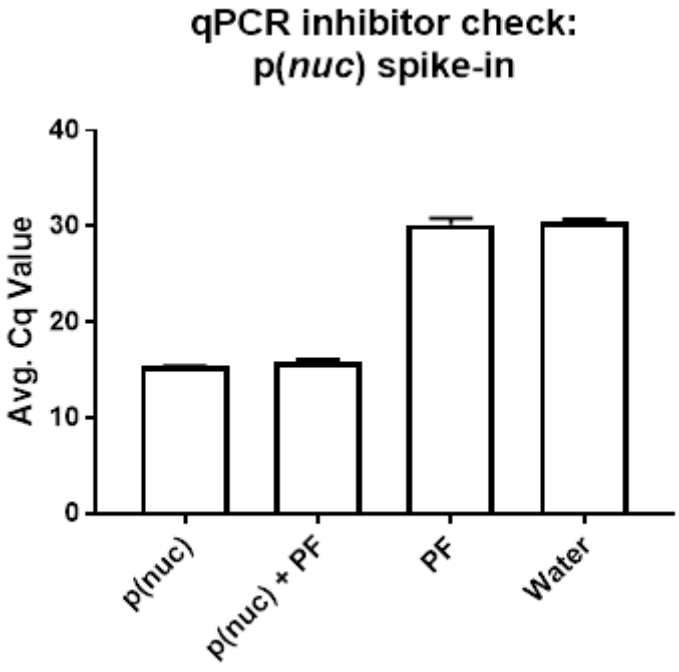


Figure 22 *– Results from qPCR test to determine impact of pleural fluid on the qPCR analysis stage using nuc primers. The pleural fluid derived template DNA was from sample 3 (see Table 20 above). Average Cq values: p(nuc) = 15.35; p(nuc)+PF = 15.78; PF = 30.19; Water = 30.54. Error bars show standard error. Raw data can be found in Appendix 9.12.1.*

The samples not containing the p(*nuc*) standard (PF and water) showed equivalent average Cq values, confirming that the pleural fluid did not contain a background level of *S. aureus* which would interfere with these experiments. For the samples containing the p(*nuc*), the average Cq values were also equivalent regardless of the presence of the extract derived from pleural fluid samples. This demonstrated that there were no significant inhibitors present in the pleural fluid DNA extract which may affect the results in the *S. aureus* specific qPCR run with the *nuc*_qPCR primers. It therefore follows that, as all components in the generic 16S rRNA run are identical, the only potential factor remaining which could affect the results is the 16S rRNA primers themselves.

5.4.3. qPCR inhibition test: 16S rRNA primers

The same procedure was carried out using the 16S rRNA primers to determine whether the efficacy of the qPCR analysis stage would be impaired by the pleural fluid when using this primer set. It was expected that the pleural fluid would contain a background level of microorganisms which would be detected by the 16S rRNA primers. Therefore, it was expected that the two sample types not containing the p(16S) standard would not produce equivalent results to each other with this primer set as they did with the *nuc* primers. Two technical replicates for each sample were analysed by qPCR using 16S rRNA primers (Figure 23; raw data in Appendix 9.12.2).

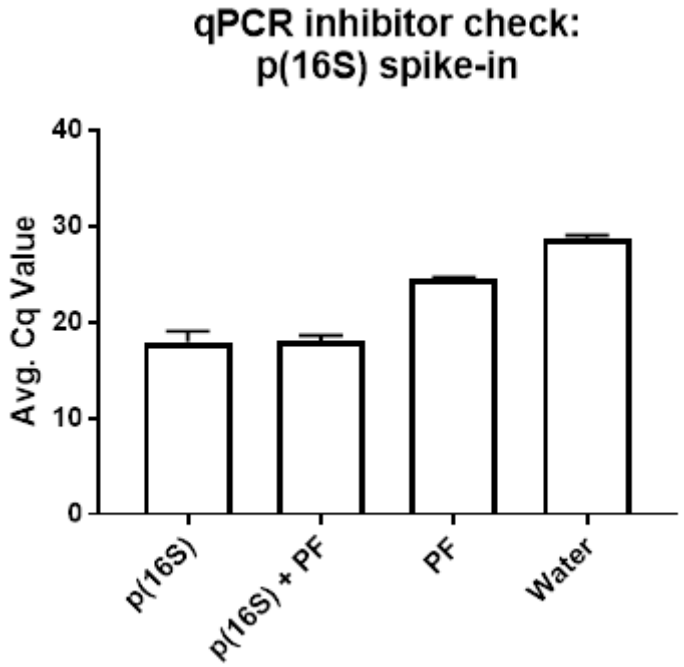


Figure 23 *– Results from qPCR test to determine impact of pleural fluid on the qPCR analysis stage using 16S rRNA primers. The pleural fluid derived template DNA was from sample 3 (see Table 20 above). Average Cq values: p(16S) = 17.92; p(16S) + PF = 18.07; PF = 24.47; Water = 28.68. Error bars show standard error. Raw data can be found in Appendix 9.12.2.*

This experiment was to determine if the signal produced by p(16S) would be masked by anything present in the clinical sample. The sample containing DNA extract derived from pleural fluid but no p(16S) standard showed the background level of microorganisms present in the pleural fluid; this sample showed a difference to the water sample as expected, but importantly it also showed a significant difference to the two samples containing p(16S). As there was a significant difference between the PF and the two p(16S) samples, it was not expected that the background level of organisms in the pleural fluid would have a significant impact on either of the p(16S) samples due to the exponential nature of qPCR. The two p(16S) containing samples showed equivalent average Cq values, this suggests that the presence of the pleural fluid had no significant impact on the efficacy of the qPCR analysis when using 16S rRNA primers.

5.4.4. Impact of pleural fluid on BB+P/C DNA extraction efficacy

Once it had been demonstrated that the pleural fluid did not contain any inhibitors which would significantly affect the qPCR analysis stage, all DNA extracts (Table 20, above) were analysed. This experiment was to determine whether the use of a complex clinical sample would impair the

efficacy of the extraction stage; this was achieved by using qPCR to compare DNA extracts either from a clinical sample or from a pure culture of *S. aureus*. As it had already been established that the pleural fluid used had no significant impact on the qPCR analysis stage with each primer set, any differences seen must be due to the extraction stage itself.

First, qPCR analysis using the *Staphylococcus aureus* specific *nuc* primers was carried out on all DNA extracts from the spiking experiments (Table 20, above). Average Cq values were taken for each sample type and plotted as a bar chart (Figure 24, raw data in Appendix 9.13.1). The qPCR analysis showed the same pattern for both the PBS and the PF sample types; both the un-spiked PBS and un-spiked PF samples showed significantly higher Cq values than the two spiked samples (PBS+JE2 and PF+JE2). The un-spiked PBS sample did have a slightly higher average Cq value than the un-spiked PF sample, this was to be expected as the clinical sample was far more complex than the blank sample of PBS and due to the exponential nature of a PCR reaction, slight differences have more of an impact at the higher Cq values.

Furthermore, the two spiked samples showed comparable Cq values, reflecting the similar quantity of *S. aureus* detected in both sample types; the similarity between the PBS+JE2 and PF+JE2 samples indicates that there was no significant loss of target material during the DNA extraction step despite the complexity of the clinical sample (PF+JE2). Although a difference was detectable between the two un-spiked samples, the lack of a significant difference between the two spiked samples shows that any impairment the clinical sample may have on the efficacy of the extraction method does not have a large enough impact to produce a false negative result from a positive sample.

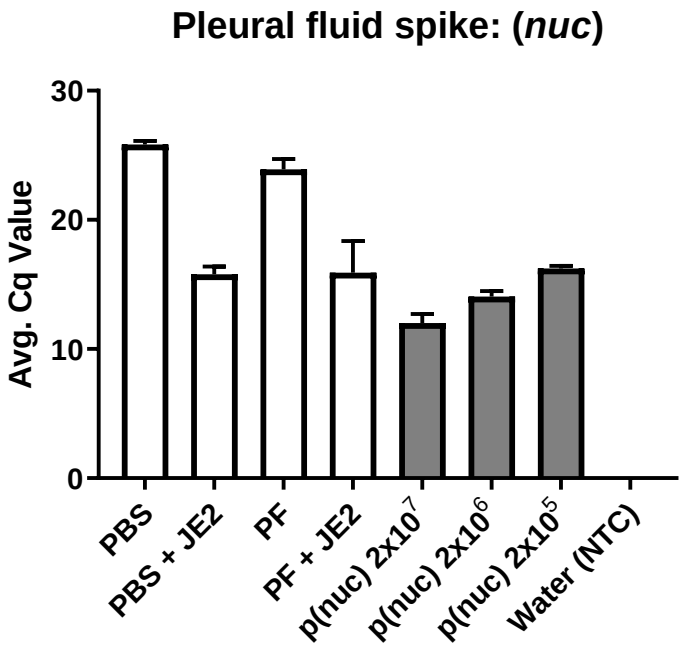


Figure 24 *– Results from qPCR test to determine impact of pleural fluid on the BB+P/C DNA extraction method when using nuc primers. Both the PBS based samples and the pleural fluid based samples showed same pattern, indicating that the presence of pleural fluid does not reduce the efficacy of the DNA extraction method or inhibit the qPCR analysis. Average Cq values: PBS = 25.84; PBS+JE2 = 15.79; PF = 23.92; PF+JE2 = 15.90; p(nuc)2x107 = 12.00; p(nuc)2x106 = 14.05; p(nuc)2x105 = 16.24; Water = 0.00. Error bars show standard error. Raw data can be found in Appendix 9.13.1*

qPCR analysis using the generic 16S rRNA primers was also carried out on these same DNA extracts (Table 20, above). Again, average Cq values were taken for each sample type and plotted as a bar chart (Figure 25; raw data in Appendix 9.13.2). The PBS samples showed the same pattern with the 16S rRNA primers as they did with the *nuc* primers, this was expected as the vast majority of organisms present in these samples should be *S. aureus* and any cells detected with the *nuc* primer set should also be detected with the 16S rRNA primers. The pleural fluid samples also followed the same pattern, however the difference between the two un-spiked samples was far greater than in the *nuc* qPCR analysis. The un-spiked PF sample was expected to show a lower Cq value than the un-spiked PBS sample due to a background level of microorganisms in the pleural fluid, this likely accounts for the greater difference in these samples when using 16S rRNA primers rather than *nuc* primers. Despite this, the two spiked samples again show very similar Cq values, showing that the composition of the clinical sample did not cause any significant loss of efficacy in the qPCR analysis stage. Regardless of appearing to be relatively similar to the samples

containing the JE2 spike, the un-spiked PF sample is significantly different to the two spiked samples due to the logarithmic nature of qPCR, with each 1 Cq difference representing a doubling of template DNA present in the sample.

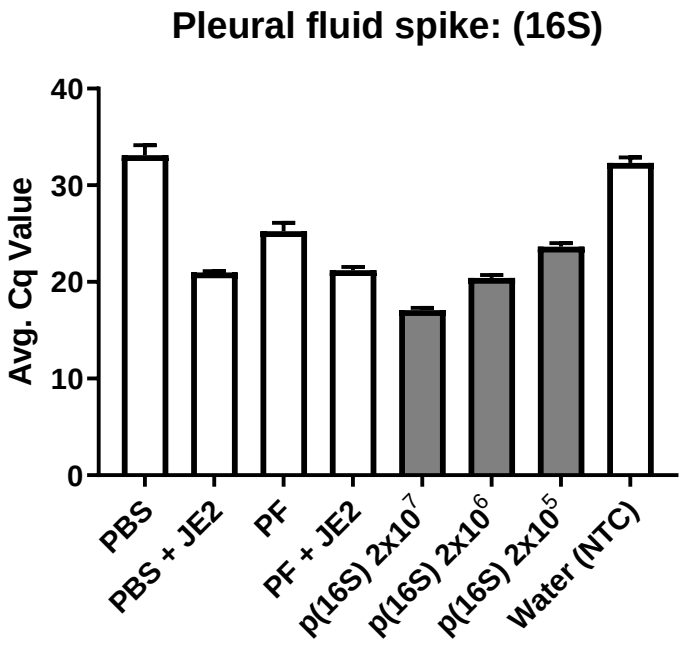


Figure 25 *– Results from qPCR test to determine impact of pleural fluid on the BB+P/C extraction method when using 16S rRNA primers. Both the PBS based samples and the pleural fluid based samples showed same pattern, indicating that the presence of pleural fluid does not reduce the efficacy of the DNA extraction method or inhibit the qPCR analysis. Average Cq values: PBS = 33.11; PBS+JE2 = 20.98; PF = 25.25; PF+JE2 = 21.22; p(16S)2x107 = 17.08; p(16S)2x106 = 20.42; p(16S)2x105 = 23.66; Water = 32.30.* Error bars show standard error. *Raw data can be found in Appendix 9.13.2.*

The spike used in these experiments was a 10µl aliquot of liquid culture of the *S. aureus* strain JE2. The culture contained $1.83x10^9$ CFU/ml, meaning the 10µl spike contained approximately $1.83x10^7$ CFU of JE2. The samples were eluted in 60µl of low-EDTA TE buffer and 1µl of this DNA extract was used for the qPCR analysis, therefore, the Cq values produced by the four test samples (PBS, PBS+JE2, PF and PF+JE2) are representative of 1/60th of the initial samples. To calculate absolute quantities of *S. aureus* in each sample, a standard curve was produced from the p(*nuc*) and p(16S) standard samples used at $2x10^7$, $2x10^6$ and $2x10^5$ copies per µl (Figure 26). Copy numbers of the *nuc* gene and 16S rRNA gene were then interpolated from this standard curve and used to calculate the number of cells thought to be in the sample (Table 21). As *S.*

aureus has a single copy of *nuc*, the copy numbers interpolated from the *nuc* standard curve were directly proportionate to the number of cells in each sample. However, *S. aureus* strain JE2 has 5 copies of the 16S rRNA gene; it was assumed that the vast majority of microorganisms present would be from the JE2 spike, so the copy numbers interpolated from the p(16S) standard curve were divided by 5 to calculate the number of cells in each sample. The spike contained ~$1.83x10^7$ cells and the DNA was resuspended in 60µl, meaning that in each 1µl of DNA extract is expected to represent ~305,000 cells ($1.83x10^7/60$).

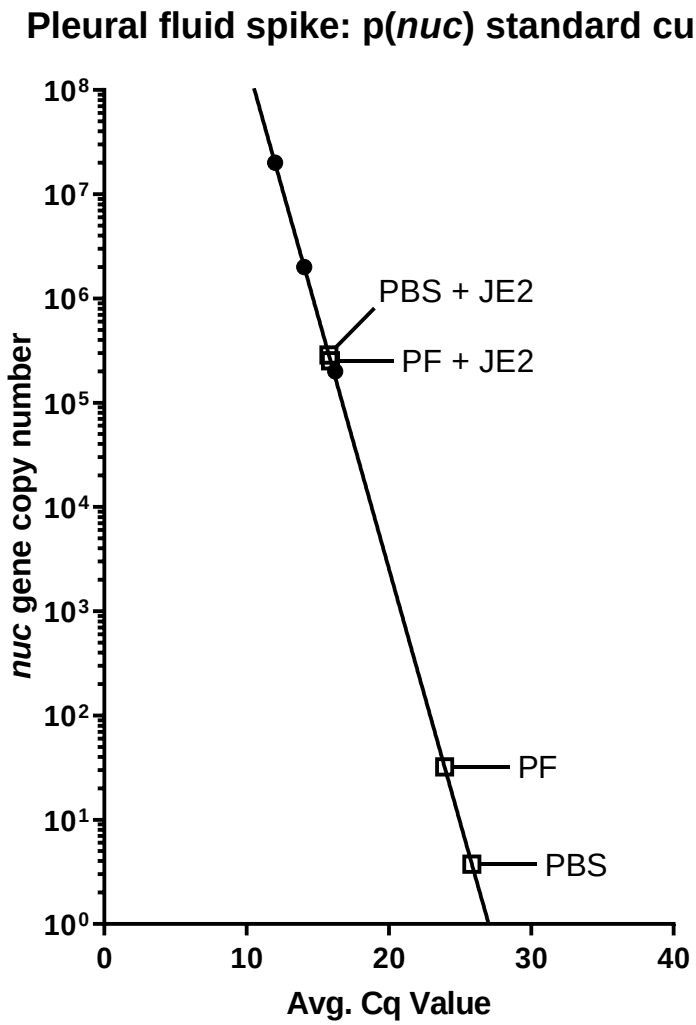


Figure 26 (A) *– Standard curves for each plasmid standard used in qPCR experiments to determine impact of pleural fluid on BB+P/C DNA extraction efficacy. A) p(nuc) standard curve; B) p(16S) standard curve. Standard curves from plasmids at a known copy number were used to interpolate the copy number from the samples with unknown copy numbers.*

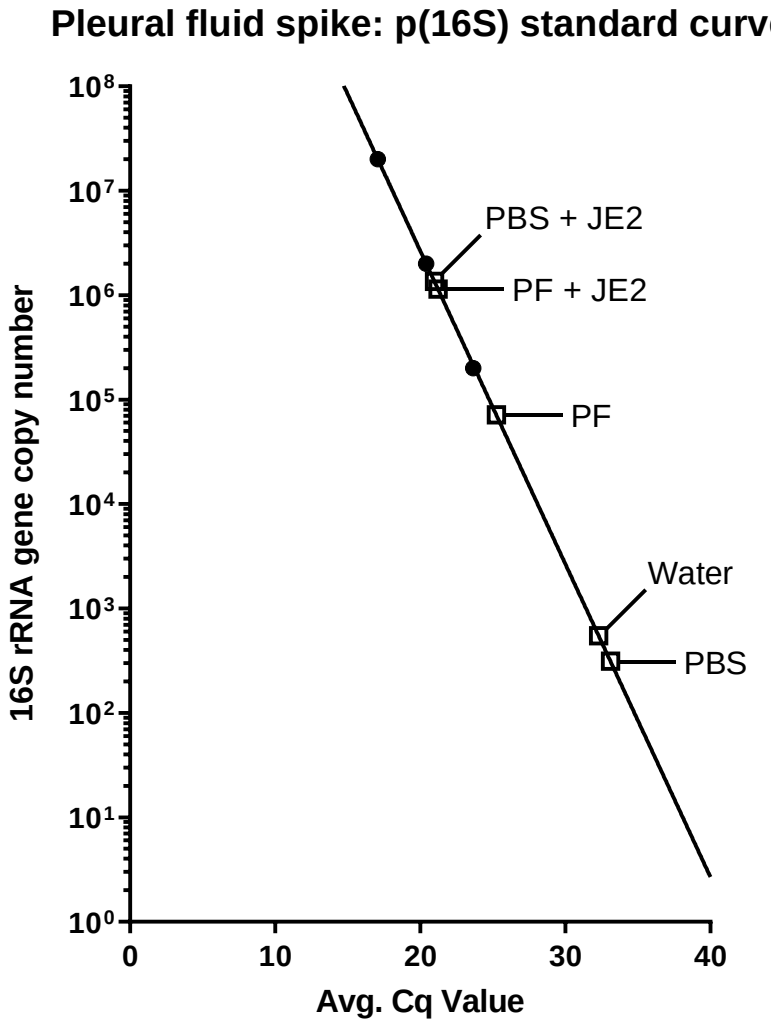


Figure 26 (B) *– Standard curves for each plasmid standard used in qPCR experiments to determine impact of pleural fluid on BB+P/C DNA extraction efficacy. A) p(nuc) standard curve; B) p(16S) standard curve. Standard curves from plasmids at a known copy number were used to interpolate the copy number from the samples with unknown copy numbers.*

	***nuc*_qPCR primers**			**16S rRNA primers**		
	Avg. Cq value	***nuc* copy no. (interpolated)**	**Number of cells**	**Avg. Cq value**	**16S copy no. (interpolated)**	**Number of cells***
PBS	25.84	3.76	4	33.11	312.28	62
PBS + JE2	15.79	286,996.40	286,996	20.98	1,356,566.34	271,313
PF	23.92	32.07	32	25.25	71,432.54	14,287
PF + JE2	15.90	253,112.90	253,113	21.22	1,146,652.84	229,331
Water (NTC)	>40	0	0	32.30	547.92	110

Table 21 *– Absolute number of cells predicted to be in the 1ul of each sample which was analysed by qPCR. Copy number determined by interpolation from standard curves (Figure). *Number of cells for 16S rRNA presumed to be S. aureus cells containing 5 copies each.*

The pleural fluid samples were anticipated to contain excessive amounts of human DNA as well as DNA from other microorganisms present in the raw clinical sample. It was hypothesised that this disproportionate amount of human DNA compared to the DNA from the JE2 spike may prevent accurate quantitation of the spike in the clinical samples. However, this experiment has demonstrated the capability of the extraction method and the specificity of the qPCR analysis to accurately detect *S. aureus* in clinical samples despite the background level of human DNA. Unfortunately, it was not possible to confirm the quantity of human DNA in the pleural fluid samples due to ethical constraints regarding the amplification of patient DNA without informed consent.

These analyses aimed to evaluate the impact that a DNA extract derived from a complex clinical sample might have on the BB+P/C DNA extraction method and its associated qPCR analysis with both *S. aureus* specific *nuc* primers and generic bacterial 16S rRNA primers. It was possible to determine the presence and absolute quantity of amplifiable DNA in the spiked and un-spiked samples for both the simplistic PBS samples and the complex clinical samples of pleural fluid, with the spiked samples showing comparable results for both primer sets. Through this comparison, it was possible to determine that there was no significant impact on the acquisition of accurate and reliable results when using the BB+P/C DNA extraction method on these complex clinical samples relative to a more simplistic sample type.

6. Discussion

6.1. Comparison of methods

Three DNA extraction methods were compared in this project, a kit-based method and two non-kit-based methods. Most DNA extractions performed in modern laboratories now utilise one of the many commercial DNA extraction kits to simplify and expedite the process of obtaining DNA from a bacterial culture. This is suitable for studies in which DNA is extracted from a pure culture, however if the composition of the sample is unknown, contamination from the 'kit-ome' can significantly impact the results of the study [54]. Furthermore, yield can be significantly influenced by the DNA extraction protocol used on complex microbial communities [55], this could have a significant impact when attempting to extract 100% of the DNA present in a sample.

Prior to the comparison of methods stage of this project, it was determined through laboratory-based experimentation that there was no significant difference in yield when using lysostaphin compared to Metapolyzyme with the Roche kit. Therefore, it was decided that MetaPolyzyme would be used for all further kit-based extractions (Met+Kit). The three methods were then directly compared, with the BB+P/C method producing the highest yield, followed by the Met+P/C method; both of which utilise phenol/chloroform for DNA purification. The Met+Kit method produced the lowest yield and uses a spin-column approach for DNA purification. Effect of sample preparation was then investigated by comparing differences in yield from fresh and frozen *S. aureus* samples. For the BB+P/C and Met+Kit methods there was no significant difference between extracting DNA from fresh samples compared to frozen samples of *S. aureus*, however there was a significant difference when using the Met+P/C method. Relative sensitivity of each extraction method compared to the other two methods was determined by completing DNA extractions on serially diluted samples. The BB+P/C method was able to extract a higher yield from more diluted samples, followed by the Met+Kit method, with the Met+P/C method having the lowest sensitivity of the three. In summary, the BB+P/C method was deemed the most effective for extracting DNA from samples with a low bacterial load when compared with the other two methods, without being significantly affected if samples were frozen before processing.

Throughout, these three DNA extraction methods were compared using a series of T-tests to determine the statistical significance of any observed differences. On reflection, it would have been more appropriate to use an ANOVA to analyse all three methods simultaneously, followed by a post-hoc test (such as the Tukey test) to provide deeper insight into any statistically significant differences. Furthermore, a multivariate ANOVA could have been used to determine statistical significance between the three methods when compared by qPCR across three different dilutions (Figure 13).

6.2. DNA extraction principles

DNA extraction can be divided into three key steps: Lysis, Purification and Precipitation [56]. The Met+Kit and the Met+P/C method rely upon enzymatic lysis of cells. MetaPolyzyme is an enzyme mix containing mutanolysin, achromopeptidase, lyticase, chitinase, lysostaphin and lysozyme; these six enzymes are hydrolytic against the cell wall, and in combination are thought to be able to lyse all Gram-positive and Gram-negative bacterial cells, as well as all yeast and fungal cells. For lysis of staphylococcal species, lysostaphin is included in this MetaPolyzyme mix. Lysostaphin has hexosaminidase, amidase and endopeptidase activity, enabling cellular lysis via cleavage of polyglycine crosslinks in the cellular wall of staphylococcal species [57, 58].

The BB+P/C method relies upon mechanical lysis of cells through physical bead-beating. The Lysing Matrix E (LME) tubes used in this method contain a mixture of 1.4mm ceramic spheres, 0.1mm silica spheres and one 4mm glass bead. The range of bead sizes allows mixed samples to be entirely homogenised, providing highly efficient cellular lysis. Cellular lysis in the BB+P/C and Met+P/C methods are also facilitated by use of CTAB, a cationic detergent which promotes instability of biofilms and cellular lysis [59] whilst maintaining the integrity of DNA [60, 61].

For DNA purification, the Met+Kit method uses a spin-column based approach. A binding buffer is added to each sample which facilitates binding of nucleic acids to glass fibres manufactured into the filter tubes which come with the kit. The sample is then subject to a series of centrifugation-based wash steps to remove cellular components, resulting in the sample only containing the DNA bound to the glass fibres in the filter tube. After purification, the elution buffer releases the DNA is

from the fibres and is collected in a microcentrifuge tube in a final centrifugation step; this step replaces more traditional DNA precipitation techniques.

DNA purification in the BB+P/C and Met+P/C methods uses phenol:chloroform:isoamyl alcohol. Water is the main solvent for the samples used, so two distinct phases form when phenol is added as it is immiscible with water. Mixing the sample thoroughly and then centrifuging allows many contaminants in the sample to partition into the phenol-containing phase, while the DNA remains in the aqueous/water phase. The DNA is then concentrated and further purified from salts in the precipitation step where the precipitated DNA is repeatedly washed with ice-cold 70% ethanol. DNA yield is enriched by use of linear polyacrylamide (LPA), which co-precipitates with DNA, increasing the amount of DNA recovered from each sample by increasing the amount of DNA which precipitates and reducing the amount of DNA lost during the ethanol wash. LPA is inert, so has no downstream effects during PCR or qPCR steps. LPA is very sensitive and is able to precipitate picogram amounts of DNA fragments larger than 20 base pairs from a sample [62], making it ideal for use with low-yield samples such as pulmonary clinical samples. Furthermore, LPA does not interfere with A260/280 readings, so does not affect DNA quantification [63].

6.3. BB+P/C DNA extraction method

The protocol for the BB+P/C DNA extraction method was originally produced by Dr Michael Cox (University of Birmingham**)** and Dr Leah Cuthbertson (Imperial College London) under the supervision of Professor Cookson and Professor Moffatt at Imperial College London, as MG-SOP-0082 (Molecular Genetics and Genomics Standard Operating Procedure). It was provided for use in this project by Dr Michael Cox.

The original protocol included preparation steps for a wide range of sample types. For this study, the most relevant were the preparation steps required for processing bacterial isolates, or pulmonary samples of bronchoalveolar lavage and pleural fluid. The preparation steps for both sample types included the addition of Hexadecyltrimethylammonium bromide (CTAB) for cellular lysis, CTAB is a cationic detergent with a structure containing quaternary ammonium compounds (QACs). The mode of action of QAC containing detergents is through electrostatic stresses

between the positive charge of the QACs and negatively charged sites on the microbial cell wall, resulting in lysis of the cell [59].

The next step in the extraction protocol is the addition of aluminium ammonium sulphate (AAS) to the sample, AAS is a flocculating agent which coagulates any dissolved organic matter which may be present in the sample [64]. As some types of organic matter may co-precipitate with DNA, the addition of AAS helps to improve the purity of the DNA after extraction. Following the addition of AAS, the sample is transferred to an Lysing-matrix E (LME) tube which contains 1.4mm ceramic spheres, 0.1mm silica spheres and one 4mm glass bead [65]. These tubes are used for the physical bead-beating of complex samples, such as tissues, and contain a range of bead sizes to maximise the amount of physical disruption.

Phenol:chloroform:isoamyl alcohol (PCI) is then added to each sample and the sample is subjected to physical bead-beating before being centrifuged. Phenol and water are immiscible, so upon addition to the sample, two phases are formed: a water-based (aqueous) phase on top of a more dense phenol-based (organic) phase. The bead-beating stage has two functions: firstly, it causes physical degradation of the cells in the sample, allowing the release of nucleic acids; and secondly, it thoroughly mixes the sample. Samples are then centrifuged, facilitating the reformation of the two phases. Mixing the sample well allows all sample components to come into contact with both of the available solvents: water and phenol. Water is a more polar solvent than phenol, and as polar compounds are generally more soluble in more polar solvents, the polar nucleic acids will remain in the aqueous phase. However, when proteins are exposed to a less polar solvent, such as phenol, their tertiary conformation changes due to the less-polar amino acids moving to the exterior of the protein, effectively denaturing it [66]. This less tightly folded state is more chemically stable for the protein but is prevented from forming in more-polar solvents, such as water. Therefore, when proteins are in a sample containing both phenol and water, they will remain in the phenol-based organic phase. When the sample is centrifuged, the phases separate into their distinct sections, also separating the proteinaceous material from the nucleic acids. Chloroform has a much higher density than water, so using a phenol:chloroform mix rather than just phenol prevents inversion of the phases [67]. This is especially useful when the exact sample composition

is unknown, such as when using clinical samples. The increased density also reduces the size of the interphase layer, reducing the opportunity to contaminate the aqueous phase when separating it from the organic phase. Isoamyl alcohol is also added as it is an anti-foaming agent [67], preventing foaming during the bead-beating process.

After the proteins have been separated from the nucleic acids in the sample, chloroform:isoamyl alcohol (CI) is added and the sample is centrifuged again. The additional chloroform added removes any remaining phenol from the aqueous layer, further purifying the sample by preventing any residual phenol from having any downstream effects on the sample. Again, isoamyl is included as an anti-foaming agent. The aqueous phase is then transferred to a new microcentrifuge tube containing a very small amount of linear polyacrylamide (LPA). At this point, the protocol provides the option to perform a double extraction. This takes the remaining sample in the LME tube and exposes it again to all previous steps of the protocol from the addition of the CTAB buffer to the removal of the aqueous phase. This can increase the yield obtained from the sample and is useful when attempting to extract as close to 100% of the DNA in the sample as possible.

The microcentrifuge tube which the aqueous phase is transferred to contains LPA, which is a co-precipitant of DNA and enables the precipitation of picogram amounts of nucleic acids in a sample [62, 68]. Being able to recover such small quantities of DNA is incredibly useful both when attempting to extract DNA from low-yield samples, and when attempting to extract 100% of the DNA present in a sample. LPA is synthetically produced which reduces the likelihood of it containing any contaminating nucleic acids compared to chemicals derived from a biological source [69]. It is also a chemically inert substance [70], so will not have an influence on any downstream applications, such as qPCR analysis of DNA extracts.

Polyethylene glycol in a salt solution (PEG/NaCl) is added to the aqueous phase of the sample. PEG is a neutral polymer which is commonly used for the precipitation of nucleic acids through salt-induced DNA condensation. The high concentration of salt present in the PEG/NaCl solution neutralises the negative charge of the phosphate backbone [71] and the PEG readily binds to substantial amounts of water, allowing the neutralised DNA molecules to aggregate [72].

After 18 hours, the sample is centrifuged and the PEG/NaCl solution removed. Following this, the pellet is washed several times with ice-cold ethanol to remove any precipitated salts remaining from the PEG/NaCl solution. After washing, the sample is left to air dry, allowing any residual ethanol to evaporate, and the pellet is resuspended in a low-EDTA TE buffer. TE buffer contains tris(hydroxymethyl)aminomethane (Tris) and ethylenediaminetetraacetic acid (EDTA); tris is a commonly used pH buffer, and EDTA chelates cations to protect the DNA sample from degradation. A low-EDTA TE buffer is used as, at high concentrations, EDTA may have an inhibitory effect on PCR by chelating the Mg^{2+} cations required for an effective PCR reaction. However, at low concentrations, EDTA can improve the efficiency of PCR by chelating any contaminating ions present as it has a relatively low affinity for magnesium compared to other metal ions which may inhibit PCR, such as Ca^{2+}, Pb^{2+}, Cd^{2+} and Hg^{2+} [73]. If present, these ions with a higher affinity to EDTA will displace the magnesium and bind to EDTA, preventing them from binding to the DNA polymerase enzyme and potentially inactivating it, reducing PCR efficiency.

This DNA extraction protocol should be carried out on ice as much as possible, and centrifugation steps should be performed in a temperature controlled centrifuge at 4°C. Keeping the sample cold throughout the purification stage of the extraction protocol reduces the activity of any DNases which may be present in the original sample, increasing the final yield of DNA. Keeping the sample cold during the precipitation stage reduces the solubility of the aggregated DNA, facilitating nucleic acid precipitation. The resuspended DNA extract is stable to be stored at -20°C or -80°C until required, however repetitive freezing and thawing is not recommended as it may damage the DNA in the sample.

6.4. Phase lock gel

After performing the BB+P/C DNA extraction method repetitively on similar types of samples, it became clear that there was some variability between technical and biological repeats of the same sample. One potential technique to reduce this variation is to use phase lock gel [47] when decanting the DNA-containing aqueous phase into a fresh tube during the extraction. After adding phenol:chloroform:isoamyl alcohol (PCI) and centrifuging the sample, it settles into two distinct phases: an aqueous phase above an organic phenol-containing phase; this occurs as phenol is

denser than water. The proteinaceous material from the cells settles in the organic phase, with some cellular debris also settling between the phases as an interphase layer. The aqueous phase contains the DNA and must be removed without disrupting the interphase as if this layer is disrupted it will contaminate the sample. This step introduces sample variation due to human error as the removal of the entire aqueous phase is confirmed by eye and is reliant on the individual's abilities with a pipette.

Phase lock gel is a viscous liquid which settles between the organic and aqueous phases after centrifugation of the sample, forming a barrier and preventing the opportunity to interrupt the interphase (Figure 27). This makes it significantly easier to decant the aqueous phase without contaminating the sample, as well as reducing the amount of DNA discarded due to human error. Furthermore, the use of phase lock gel provides increased protection from the phenol contained in the organic phase, making the procedure less hazardous to the health of the individual performing the extraction.

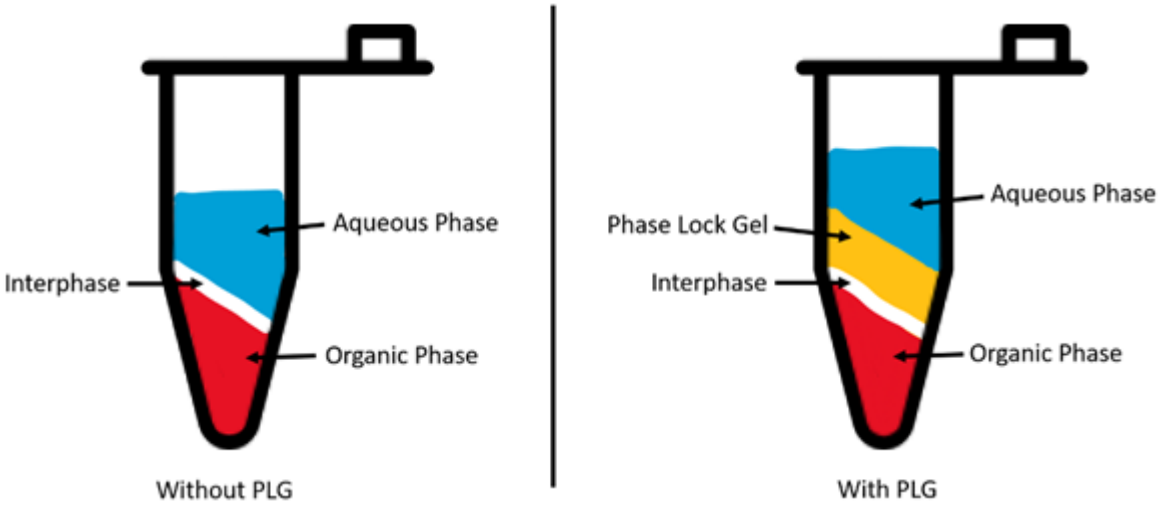


Figure 27 *– Diagram showing how addition of phase lock gel (PLG) forms a barrier between the aqueous and organic phases, allowing removal of the DNA-containing aqueous phase without interrupting the interphase layer of protein and cell debris.*

The experiments which were performed to assess the impact of phase lock gel on the BB+P/C method demonstrated that it is an effective method to reduce the variation in yield between samples, and also had the further benefit of reducing the loss of sample material during the DNA extraction process. One caveat with the experimentation performed to assess the impact of phase lock gel is the use of the NanoDrop spectrophotometer as the sole method of DNA yield

quantification as there was no way of identifying any factors within the samples which may cause interference with the A260/280 reading. Confidence in the reading obtained was increased by taking three independent readings per sample, however if the whole sample was contaminated this provided no explanation as to why yield may have been affected. Interestingly, the results showed unequal variance for the samples with PLG compared to the samples without PLG; therefore for these samples the parameters of the t-test carried out were set to assume unequal variance. Following these experiments, phase lock gel was implemented in all BB+P/C DNA extractions performed in order to reduce artificial sample variability.

The phase lock gel used in these experiments was produced in-house due to the high cost of buying it commercially. The phase lock gel produced was suitable for the requirements of this project, however after a period of 6 months, the gel would begin to separate back into its constituent components. Therefore, for a longitudinal study the recommendation from this project would be to purchase the more costly, but quality controlled and more stable, commercial phase lock gel.

6.5. Limit of detection (LoD): Extended serial dilution experiments

To find the limit of detection (LoD) of the BB+P/C DNA extraction method and the associated qPCR analysis, DNA was extracted from a serial dilution series of *S. aureus* strain JE2. Two different primer sets (*nuc*_qPCR and 16S rRNA) were used, with an independent LoD determined for each set.

The qPCR analysis performed on the DNA extracts using the *nuc_qPCR* primers showed that this method is able to extract and detect *S. aureus* cells in a 1ml sample down to a level of 1,000 cells. This is almost the lowest dilution that would be possible to detect under the conditions used; this is because each 1ml sample produced a DNA extract of 60µl, and only 2µl of the total 60µl DNA extract was used in the qPCR analysis stage. This 2µl represents only $1/30^{th}$ of the original sample; the theoretical number of cells represented in qPCR analysis for each dilution sample used in these experiments is shown in Table 22.

To demonstrate the issue, imagine a hypothetical 1ml sample containing 100 copies of a gene; assuming perfect extraction efficiency, the 60µl DNA extract produced would contain all 100 copies

of the gene. The 2µl which is then taken for qPCR analysis would theoretically only contain 3 copies of the gene. Another hypothetical sample containing only 10 copies of a gene, treated in the same way, would theoretically have 0.3 copies in the 2µl used for qPCR analysis; probability dictates that this sample would usually not produce a result despite the DNA extract containing 10 copies of the gene. Therefore, even if the extraction efficiency is 100%, not all organisms in the sample will be detected in qPCR when processed under these conditions as the whole DNA extract is not analysed. This must be considered if attempting to calculate absolute values from the Cq values produced and it may be preferable in future studies to resuspend the DNA in a lower volume and analyse the entire DNA extract in the qPCR run. This has the caveat of not permitting technical repeats of a sample at any stage, however biological repeats are possible to maintain the scientific integrity of a study. Samples were not processed in this way for this project as technical repeats from the same sample were essential for investigating the fundamental abilities of this DNA extraction method and qPCR analysis.

The qPCR analysis performed found the LoD cut-off for a 1ml sample to be 1,000 cells when using the *nuc* primers. Following the extraction and resuspension in 60µl, the 2µl of this sample that was used for qPCR theoretically contained ~33 copies of the *nuc* gene (Table 22). This accounts for the difference in LoD seen between these JE2 samples (*nuc* LoD = $1x10^3$ cells) and the p(*nuc*) samples which were found to have an LoD of 20 copies. As expected, the 16S rRNA primers could not reliably detect the *S. aureus* cells to the same extent as the *nuc* primers, likely due to being less specific to their target. 16S rRNA analysis was able to reliably detect down to a dilution of 10,000 cells in 1ml; the 2µl from this sample used for qPCR analysis theoretically represented ~333 cells (Table 22). Again, by accounting for the actual number of cells represented in the qPCR analysis stage, the LoD for the extracted samples (16S LoD = $1x10^4$) is comparable with the LoD for the p(16S) standard which was found to have an LoD of 200 copies.

Sample	Theoretical CFU of *S. aureus* JE2 in 60µl DNA extract	Relative number of cells represented in 2µl used for qPCR analysis (CFU/30)
1	1×10^9	*Not used for qPCR
2	1×10^8	3,333,333
3	1×10^7	333,333
4	1×10^6	33,333
5	1×10^5	3,333
6	1×10^4	333
7	1×10^3	33
8	100	3
9	10	0
10	1	0
Blank	0	0

***Table 22** – Number of cells represented in qPCR for each dilution used in LoD experiments. Total DNA was resuspended in 60µl and only 2µl of this was used for qPCR. *sample not used for qPCR due to excessive amount of DNA which would affect acquisition of data for all samples.*

This difference between the number of copies present in the total 60µl DNA extract compared to the number of copies in the 2µl analysed by qPCR is the main factor accountable for the difference in LoD between the plasmid standards (p(*nuc)* = 20 copies, p(16S) = 200 copies) and extracted samples of JE2 (*nuc* = $1x10^3$ cells, 16S = $1x10^4$ cells). Following the BB+P/C extraction, the DNA is resuspended in 60µl in order to produce enough DNA to perform any experiments with multiple repeats if necessary, however this volume could theoretically be reduced in future studies and LoD scores recalculated if the scores determined by this project are not sensitive enough for the intended use. The LoD of the BB+P/C method will be dependent on the exact conditions used, but has been demonstrated in this project to be at least sensitive enough to detect $1x10^3$ cells in 1ml of sample using the nuc primers and $1x10^4$ cells when using the 16S rRNA primers.

The method developed in this project is intended for use on clinical pulmonary samples which are expected to contain far higher quantities of microorganisms than 1,000 cells per ml; the clinical threshold for a positive BAL sample is currently $1x10^4$ CFU/ml [17], which is ten-fold higher than the LoD of this method when using *nuc* primers for qPCR analysis and equal to the LoD when using 16S rRNA primers. Furthermore, the starting volume of clinical samples used will not be limited to a maximum of 1ml, meaning that it may be possible to detect far lower quantities of microorganisms per ml by increasing the initial volume of sample used. For future work using this method, it will be crucial to record the starting volume of each sample used, the volume of buffer

used for resuspension of DNA, and the volume of DNA extract used for qPCR analysis. All three of these variable factors will impact the LoD of the method and are also vital to consider when calculating absolute quantities of organisms in a sample against a standard curve.

6.6. Clinical samples and pleural fluid spike experiments

The purpose of the experiments using pleural fluid was to determine whether the use of complex clinical samples would have an impact on the results acquired through the BB+P/C DNA extraction method. This was important to do as the extraction method being developed in this project is ultimately intended for use on clinical samples. The first set of experiments using pleural fluid determined that the qPCR analysis stage was not significantly impaired through use of a DNA template derived from clinical pleural fluid samples. The second set of experiments was an extension of this to determine the effects of pleural fluid on the entire BB+P/C DNA extraction and qPCR analysis process combined. These experiments found that the results were not significantly different between the simplistic samples and the clinical samples containing DNA derived from pleural fluid; therefore, it was concluded that the efficacy of the process was not significantly impaired when processing the clinical samples.

Pleural fluid was used for these experiments as it was not possible to obtain any bronchoalveolar lavage (BAL) samples. The samples were required for use during the summer months and, as the incidence rate of pneumonia is higher during the winter [1], there were very few patients who required a bronchoalveolar lavage. Furthermore, the patients with chronic disease who require a bronchoalveolar lavage more regularly are often unconscious and therefore are unable to provide consent for the use of their lavage fluid in research. Therefore, it was far easier to obtain ethically approved samples of pleural fluid compared to bronchoalveolar lavage. The patient who provided the pleural fluid used in this study required pleurocentesis due to lymphoma.

The pleural fluid used was a suitable replacement for BAL as it was an exudate which had originated within the lungs and entered the pleural cavity due to the inflammation caused by the lymphoma. As the pleural effusion was due to lymphoma, rather than due to an active lung infection, the samples were far more valuable for assessing the efficacy of this DNA extraction method. This is because the microbial load of this pleural fluid would be far lower than if the

sample was obtained from a patient with an active infection. However, it is important to note that the exact composition of a clinical sample may differ significantly between patients, and this specific pleural fluid sample which was used may not be representative of all pleural fluid samples. Therefore, any future work using this DNA extraction method on clinical samples should consider this and it is recommended that they independently verify the efficacy of the process with their specific samples if possible.

The use of a more complex sample type, such as pleural fluid, had potential to impair the DNA extraction process at either of the two key stages: the DNA extraction itself, or the qPCR analysis of the DNA extracts. Any loss of efficacy in the clinical samples compared to the PBS samples at the extraction stage was likely to be due to components in the sample which could disrupt the individual steps of the DNA extraction protocol by altering characteristics such as pH or density. Density is important as a change could result in inversion of the sample with the aqueous phase locating below the organic phase, making it difficult to recover the DNA without contaminating the sample. Furthermore, the sample must remain between pH 7 and pH 8 in order for phenol-chloroform based extractions to work; if pH decreases, the negative charges of the DNA phosphate backbone would be neutralised by the increased concentration of H+ ions, allowing the DNA to dissolve into the organic phase rather than the aqueous phase. This would prevent the isolation of DNA from the proteinaceous material present in the sample. Fortunately, the pH of pleural fluid is typically between pH 7.60 and pH 7.66 [74], so should not impact the DNA extraction in most cases. Loss of quality at the qPCR analysis stage would likely be due to either the presence of PCR inhibitors, or the substantial amount of human DNA present in clinical samples.

The first set of experiments examined the impact of pleural fluid on the qPCR analysis stage. These experiments consisted of comparisons between two sets of samples, one of which was spiked with a DNA extract derived from a clinical sample of pleural fluid. By comparing the Cq results produced from qPCR on these samples against samples not containing the pleural fluid derived DNA spike, it was possible to confirm that the pleural fluid used in this experiment did not impair the qPCR analysis stage. This was the case for both the *S. aureus* specific *nuc* primers and for the generic 16S rRNA primers. Furthermore, the first qPCR experiment using *nuc* primers

confirmed that the pleural fluid used did not contain a background level of *S. aureus* which would interfere with the results from these analyses. Pipetting error was identified as a potential source of inaccuracy in this qPCR stage as only 1µl of DNA extract was added to the qPCR reaction; variation due to pipetting error could be diminished by using a minimum of 2µl of DNA template in qPCR analysis for any future studies.

The second set of experiments examined the impact of pleural fluid on the BB+P/C DNA extraction stage by comparing DNA extracts from two samples of pleural fluid against two samples of PBS; for each sample type, one was spiked with *S. aureus* strain JE2. The similarity in Cq values for the spiked samples in both sample types demonstrated that the detection of *S. aureus* was not significantly impaired in the pleural fluid derived samples compared to the PBS samples. The similarity in Cq values using *nuc* primers for the un-spiked samples showed that there was no significant background level of *S. aureus* in the pleural fluid; however the difference in Cq values when using the 16S rRNA primers suggested that there was a background level of other microorganisms in the pleural fluid. This was expected as the pleural fluid was likely to contain a low level of organisms from the patient's natural pulmonary microbiome.

A major concern prior to this experiment was that the number of human cells present in the sample would provide an overwhelming amount of human DNA, significantly inhibiting the detection of bacterial DNA in the sample. As the pleural fluid was obtained from a patient with lymphoma it was likely to have a higher proportion of human cells (compared to bacterial cells) than if it was from a pneumonia patient with an active microbial infection. Therefore, the pleural fluid used represents a scenario with a higher than usual number of human cells present in the sample; despite this, the efficacy of the BB+P/C DNA extraction and qPCR analysis process was not significantly inhibited and it can be understood that efficacy would also not be impacted in a sample of pleural fluid from a patient with an active infection.

A serial dilution of the p(*nuc*) and p(16S) plasmid standards were also run alongside the samples in these experiments, allowing the production of a standard curve with known copy numbers for each gene. This allowed absolute quantification of the number of cells in each sample through

interpolation of copy number from the average Cq value for each sample. Unfortunately, there were several caveats associated with the absolute quantification performed in these experiments.

The first is that it was not possible to definitively establish the accuracy of the absolute quantification performed. The JE2 liquid culture used was at an estimated $1.83x10^9$ CFU/ml, however this was calculated through performing a serial dilution and plating out each dilution, and then counting the colonies on the $1x10^{-6}$ to $1x10^{-9}$ plates and extrapolating up to give the approximate CFU/ml of the neat culture. This provided an opportunity for any error due to pipetting to be magnified when scaled up from the lower dilutions to the neat culture.

Furthermore, only 10µl of the neat culture was added to each sample as a $1.83x10^7$ spike, again, pipetting error could add significant variability to the number of cells added to each sample. Combined, this reduced the confidence in the value designating the expected number of cells in each sample (~305,000 cells). The accuracy of the absolute number of cells determined to be in each sample may also be impaired by pipetting errors as of the total 60µl DNA extract, only 1µl was analysed by qPCR; this error could be reduced in future studies by analysing more than 1µl in the qPCR stage. The number of cells detected in each sample showed that the method was unable to extract and detect 100% of the cells expected to be in the sample, however it was impossible to know the exact number of cells actually present in the sample, and therefore the absolute quantification for each sample can only be used as a rough estimate when compared to the expected number of cells in the samples. It is useful, however, for comparisons between different samples of the same type, such as the spiked vs. un-spiked samples analysed with each primer set.

In addition to this, the absolute quantification of the samples analysed with 16S rRNA primers is impacted by the variation in 16S copy number between species. Correcting for 16S rRNA gene copy number in microbiome studies remains an unsolved problem [75]; *S. aureus* strain JE2 has 5 copies of the 16S rRNA gene in its genome [76], however this number can vary greatly between different microorganisms with some having just one copy, and others having as many as 21 copies (e.g. *Photobacterium damselae* [77]).

To quantify the number of JE2 cells in the sample in absolute terms, it was assumed that all cells present in the sample were *S. aureus* JE2, however this is highly unlikely to be the case for the pleural fluid derived samples. The absolute copy number detected through qPCR was divided by 5 to determine the number of JE2 cells present in the sample, however if there were organisms with different 16S rRNA copy numbers, these would be either over- or under-represented in the absolute number of cells quantified in these samples.

It was calculated that the un-spiked sample of pleural fluid contained approximately $1.4x10^4$ cells. Each pleural fluid sample was 5ml, meaning the pleural fluid contained an estimated $2.8x10^3$ cells per ml. There is a possibility that the quantity of cells in the un-spiked pleural fluid increased due to cellular growth when transporting the pleural fluid to the University of Bristol or when defrosting it at room temperature as human pleural fluid has been shown to be a potent growth media for some bacteria [78].

It is also important to note that these experiments did not prove the absence of inhibitors in the pleural fluid sample, only that anything present in these clinical samples had no significant impairment on the BB+P/C DNA extraction method and the associated qPCR analyses. It is still possible that the DNA extract produced contains various inhibitory components which may impact downstream applications, such as if the DNA extract is used for sequencing. Furthermore, every clinical sample is unique to each individual patient and therefore may show inherent differences in its composition. Therefore, the clinical samples used in future studies may contain inhibiting components which were not present in this study, however, despite any differences between clinical samples, they will also have many similarities. This study demonstrates that the DNA extraction and qPCR analysis was not significantly impaired by using this pleural fluid, and therefore it can be concluded that any components common to all pleural fluid samples (and were therefore present in the sample used) do not impact the results acquired through the BB+P/C DNA extraction method.

Another factor to consider in future projects concerns the viability of the organisms. This method did not include any steps to remove any extracellular DNA present in the sample prior to the DNA extraction. Therefore, the method is unable to distinguish whether the organisms present in the

pleural fluid were viable at the start of the BB+P/C DNA extraction procedure. Any future work must consider their objectives with regard to the viability of organisms within the sample and carry out appropriate steps to remove any extracellular DNA prior to starting the BB+P/C DNA extraction if only viable organisms are desired.

6.7. PCR Primers

6.7.1. *nuc* primers

The *nuc* gene encodes a *Staphylococcus aureus* specific thermonuclease [79]. Primers for this gene were chosen as it is well established as a PCR target for the specific identification of *S. aureus* [80-82], allowing *S. aureus* presence or absence in a sample to be determined without the need to sequence the extracted DNA**.** The initial *nuc* primer set used in this project was the *nuc*_Hoegh set. The sequence for these primers was taken from literature and were originally designed by a group at the Department of Clinical Microbiology in Odense University Hospital, Denmark [40]. They were produced for a project interested in reliable identification of *S. aureus* in clinical microbiological diagnostics, so appeared to be a good primer set to use for this project. Unfortunately, it was later discovered that this primer set produced a product with *Streptococcus pneumoniae* strain d39. This was not a concern for the study carried out by Hoegh *et al.* as the samples used in their study were clinical isolates which had been phenotypically identified as *S. aureus*, and therefore there was no possibility of another organism being detected by their *nuc* specific primer set in their experiments. It was also not a concern for the first stage of this project in which an extraction method was selected based on DNA extractions from pure cultures of *S. aureus*; however it was evident that this primer set would not be appropriate for the microbiome-based studies that the extraction method developed in this project is intended for.

To overcome this obstacle, a new primer set was designed for the *S. aureus nuc* gene. To produce a primer set that would be suitable for all *S. aureus* strains, multiple sequence alignment was performed using sequences for the *nuc* gene from 31 representative strains of *S. aureus*. These strains were selected from a list of strains used by AureoWiki to compile a pangenome for *S. aureus* [52]. From the multiple sequence alignment it was possible to determine the most homologous region of the *nuc* gene; this conserved region was selected as a target for the primers

in order to increase the likelihood of the primer set being suitable for all *S. aureus* strains. The primers designed for this region were *nuc*_qPCR_F and *nuc*_qPCR_R and produce a 180bp amplicon.

To assess the *S. aureus* specificity of these primers, the sequences were entered into the NCBI Primer BLAST tool [53]. In order to get the broadest coverage, the 'nr database' was selected and no organism was specified in the primer pair specificity checking parameters section. Of the 688 results generated, 682 were *S. aureus* strains; the only results to not match an *S. aureus* strain were 2 cloning vectors (pSEC-Nuc-OmpH and pFUN), a synthetic construct (pS120+), and 3 sequences from the very closely related *Staphylococcus schweitzeri* which is a non-human primate-associated organism. *S. schweitzeri* is phenotypically very similar to *S. aureus* and has a near identical 16S rRNA gene sequence [83] so may not distinguished from *S .aureus* in most microbiome based studies. All products produced in silico were 180bp in length, matching the expected amplicon size; this includes the 6 results which did not match an *S. aureus* strain, suggesting that the vectors and synthetic construct identified are likely to contain a copy of the *S. aureus nuc* gene.

After showing that the primers were, in theory, specific only to the *S. aureus nuc* gene, it was important to ensure the primer set would work in vitro. To do this, a qPCR experiment was performed using the *nuc*_qPCR primers on genomic DNA (gDNA) extracts from the *S. aureus* strain JE2 and the *Streptococcus pneumoniae* d39 strain which had produced a product with the *nuc*_Hoegh primer set. This experiment clearly demonstrated that a product was produced with the *S. aureus* JE2 template, but not with the *S. pneumoniae* d39 template. Together with the in-silico Primer BLAST test, these results show that these *nuc*_qPCR primers are suitable for specifically detecting *S. aureus* in a sample. It must be noted that although the primers have been shown to only produce a product with *S. aureus*, clinical samples may contain organisms not included in the NCBI database which could produce a product with these primers. Although this is unlikely, it means that there cannot be 100% certainty that the product is from *S. aureus*. To make the primer set even more specific to the *S. aureus nuc* gene, a TaqMan probe could be produced for use in

future experiments, alternatively qPCR products could be sequenced to ensure a direct match with the *S. aureus nuc* gene.

6.7.2. 16S rRNA primers

Utilisation of the 16S rRNA gene is the most common genetic-based method for generic bacterial identification. This gene is universal to all bacteria as it encodes an essential ribosomal subunit. The majority of the 16S rRNA gene is highly conserved, however it contains nine hypervariable regions which can be used to identify bacteria based on species-specific sequences. A major benefit of using the 16S rRNA gene for bacterial identification is that it bypasses the need to culture the organism in vitro for phenotypic identification; this saves a significant amount of time and resources for microbiome studies, it is also more accurate than culture based methods as it is able to identify organisms which have never been successfully cultured *in vitro* [84]. Whole 16S rRNA gene sequencing can be prohibitively expensive, so to reduce cost, only the fourth variable region (V4) was chosen for this study which is 254 base pairs in length in most bacteria, only deviating by a few base pairs between organisms [85].

V4 is a semi-conserved hypervariable region which can provide phylum level resolution as accurately as the full 16S rRNA gene [86], however primer design can introduce biases by selecting for or against certain organisms [87]. Mitigating these potential biases requires a primer set which incorporates degenerate bases in the sequences of each primer to compensate for the target DNA not being fully conserved between organisms. The degenerate primers ensure that almost all bacterial species will be captured, however the presence of degenerate bases significantly reduce the melting temperature (T_m), and therefore the annealing temperature (T_a), of the primer pair as it is determined by the lowest possible combination found in the degenerate primer set.

6.8. PCR

The degenerate nature of the 16S rRNA primers led to this primer pair having a very low annealing temperature which can have an effect on the efficiency of the PCR reaction This may explain why after agarose gel electrophoresis of PCR products, the bands were brighter for samples amplified using *nuc* primers than for samples amplified using 16S rRNA primers. Traditional PCR was able

to determine presence or absence of DNA from a sample; however, it did not offer the specificity required to determine the quantity of a gene in the sample.

The use of qPCR in this project enabled the quantification of gene products, this works through the incorporation and subsequent detection of fluorescent markers attached to the dNTP's in the reaction mix. In order to determine an absolute quantity of DNA in a sample, a standard is needed of a known amount of the gene target for each primer set. In this project, a plasmid containing the *nuc* gene from *S. aureus* strain JE2 (p(nuc)) was produced to be used with the *nuc* primer set, and a plasmid containing the full 16S rRNA gene from *Vibrio natriegens* (p(16S)) was used for the 16S rRNA primer set. Both plasmid standards were used at a quantity of $2x10^7$ copies per µl.

The 16S rRNA gene from *Vibrio natriegens* was selected as this species is not found in the human lung; it was necessary to choose a standard from an organism which is not typically found in the lung microbiome to remove the possibility of false positives due to cross-contamination from the standard. The use of plasmid standards at known copy numbers also provided the opportunity for absolute quantification of the number of copies of a gene present in a sample; this required a serial dilution series of the plasmid standard to be run alongside the samples in the qPCR run, production of a standard curve from the Cq results and interpolation of copy number from the Cq values of the unknown samples.

One potential improvement would be the use of digital PCR (dPCR) instead of qPCR. dPCR is another quantitative PCR method which has significantly increased sensitivity due to the method by which the sample target is measured. dPCR works by partitioning a sample into a vast number of individual picolitre sized droplets with either zero or one target molecules contained within each before running the thermal cycles. This results in each droplet carrying out a separate reaction. After the reaction has taken place, the droplets are analysed and counted as either positive or negative, depending on the presence or absence of amplified product in the droplet.

The contents of the sample can then be quantified based on the proportion of positive to negative droplets measured. One of the most recent advancements in dPCR technology is the ThunderBolts dPCR system by RainDance technologies which is capable of sequencing the PCR product directly

after amplification [88]. This provides the opportunity to determine the precise composition of a sample and simultaneously determine the quantity of each organism present.

7. Conclusions

During this project, three DNA extraction methods were directly compared to determine their ability to obtain DNA from low yield samples. The three methods were:

1. Met+Kit: uses MetaPolyzyme for cellular lysis and the Roche High Pure PCR Template Preparation Kit for DNA isolation and purification.
2. BB+P/C: uses physical bead-beating and CTAB detergent for cellular lysis, and traditional phenol/chloroform treatment for DNA isolation and purification
3. Met+P/C: a combination of the two above methods, using MetaPolyzyme for cellular lysis and traditional phenol/chloroform treatment for DNA isolation and purification

The first experiments performed determined the ability of the Roche kit to use MetaPolyzyme for cellular lysis with the same efficiency as Lysostaphin; these experiments showed no significant difference in yield between the two enzymatic treatments. Therefore, MetaPolyzyme was substituted for Lysostaphin for all other experiments using the Roche kit.

The three methods were then directly compared for their ability to extract DNA from samples of *S. aureus* JE2 liquid culture. The BB+P/C method produced the greatest yield, followed by the Met+P/C method, with the Met+Kit method producing the lowest yield.

As clinical samples are often frozen before treatment, each of the methods was assessed for their ability to extract DNA from fresh and frozen samples of *S. aureus* JE2 liquid culture. The BB+P/C method showed the least variation in yield between fresh and frozen samples, followed by the Met+Kit method, with the Met+P/C method showing the greatest difference between fresh and frozen samples. Therefore, the BB+P/C method was determined to be the most reliable if the state of the samples cannot be controlled.

The methods were then compared for their ability to extract DNA from low yield samples, this was investigated through extracting from a serial dilution of *S. aureus* JE2 liquid culture from neat to $1x10^{-2}$. Initial comparison through PCR was impaired by the sensitivity of the NanoDrop spectrophotometer to provide reliable quantification of the DNA yields at the lowest dilutions. Therefore, DNA extracts were compared by qPCR. This found the the BB+P/C method produced

the highest yield at all dilutions, most importantly showing a significantly higher yield than the other two methods at the $1x10^{-2}$ dilution. At this dilution, the Met+Kit produced the next highest yield, followed by the Met+P/C method.

Through these comparisons, the BB+P/C technique was shown to be the most reliable and effective of the three methods for extracting DNA from low yield samples and was selected to be more rigorously investigated and developed for use on clinical samples of bronchoalveolar lavage and pleural fluid.

To determine the limit of detection (LoD) for the BB+P/C DNA extraction method and the qPCR analysis which would be used to obtain results, the qPCR analysis stage was first investigated in isolation using the plasmid standards, p(*nuc*) and p(16S). This showed the qPCR stage to have an LoD of 20 copies with the *nuc* primers and 200 copies with the 16S rRNA primers.

To determine the LoD of the BB+P/C method and qPCR analysis combined, an extended serial dilution was performed on a liquid culture of *S. aureus* JE2. DNA was extracted from 1ml samples containing a range of *S aureus* JE2 from $1x10^{8}$ CFU/ml to 1CFU/ml and analysed by qPCR. This found the LoD to be $1x10^{3}$ CFU in 1ml with the *nuc* primers and $1x10^{4}$ CFU in 1ml with the 16S rRNA primers. The LoD was likely lower for the two stages combined than it was for the qPCR stage in isolation due to only using a fraction of the total DNA extract produced for each sample in the qPCR analysis performed.

Once the LoD of the BB+P/C method had been determined, experiments were performed to assess the impact, if any, that complex clinical samples could have on the efficacy and accuracy of the method. Again, the qPCR stage was assessed in isolation prior to investigating the complete method. For these experiments, a sample of pleural fluid was obtained from a patient at Southmead hospital in Bristol. The qPCR inhibition experiments determined that the use of a DNA template derived from the clinical sample had no significant impact with either set of primers (*nuc* and 16S rRNA), these experiments also established that the pleural fluid used did not contain a background level of *S. aureus* and was therefore suitable for the following experiments to determine the efficacy of the complete BB+P/C method.

To determine the impact of the clinical sample of the efficacy of the BB+P/C method and associated qPCR analysis combined, the pleural fluid was spiked with *S. aureus* JE2 prior to the extraction and the DNA extracts were compared against spiked samples of PBS which were processed in parallel. This experiment found no significant difference on the amount of *S. aureus* JE2 detected in the clinical samples compared to the PBS samples, showing that the BB+P/C method is a proficient method to use for the extraction microbial DNA from clinical samples.

A dilution series of the relevant plasmid standard was also included in each qPCR run for these samples, this permitted absolute quantification of the number of cells in each initial sample. This was successful and could be very useful for future studies using this method, however there were several caveats recognised. The first is that only a fraction of each sample was run in qPCR, so results were statistically revised to determine the number of cells in the initial sample, this provides an opportunity for any error to be magnified after qPCR analysis. There were also issues in determining the accuracy of the absolute quantification step as the precise number of cells expected to be in each sample was unknown. Finally, there were issues with calculating the number of cells in the samples analysed with the 16S rRNA primers as different microorganisms have variable copy numbers of the 16S rRNA gene, producing difficulty when converting from copy number in a sample to the number of cells present.

Ultimately, this project resulted in an extraction method capable of extracting microbial DNA from low yield clinical samples. The method is not impacted by whether the sample is fresh or frozen and is highly reliable, with very little variation between results produced from identical samples. Each primer set used for qPCR also has an associated plasmid standard which can be used for absolute quantification of the number of cells in the initial samples used.

8. Future directions

There is scope for further work into the development of this DNA extraction protocol. One suggested investigation would be to perform the BB+P/C DNA extraction on a microbial community standard, such as the ones available from ZymoBiomics [89]. Using a mock community of known quantities such as this would allow direct comparison between the amount of each organism present in the sample before the extraction against the quantity detected by qPCR analysis. Being a diverse, mixed community, this would also allow comparison of extraction efficiencies between different types of microorganisms (for example, Gram-positive vs Gram-negative). This could be completed by sequencing the DNA extracts using the 16S rRNA primers, followed by 16S microbial analysis which could be carried out using pre-existing pipelines, such as Mothur [90]. This approach could also be used in future studies in order to determine the proportion of a specific organism within the sample, however it should be noted that correcting for 16S rRNA gene copy number remains an unsolved problem in microbiome studies [75]. These mock communities would also be useful for confirming the accuracy of the absolute quantification described in this project.

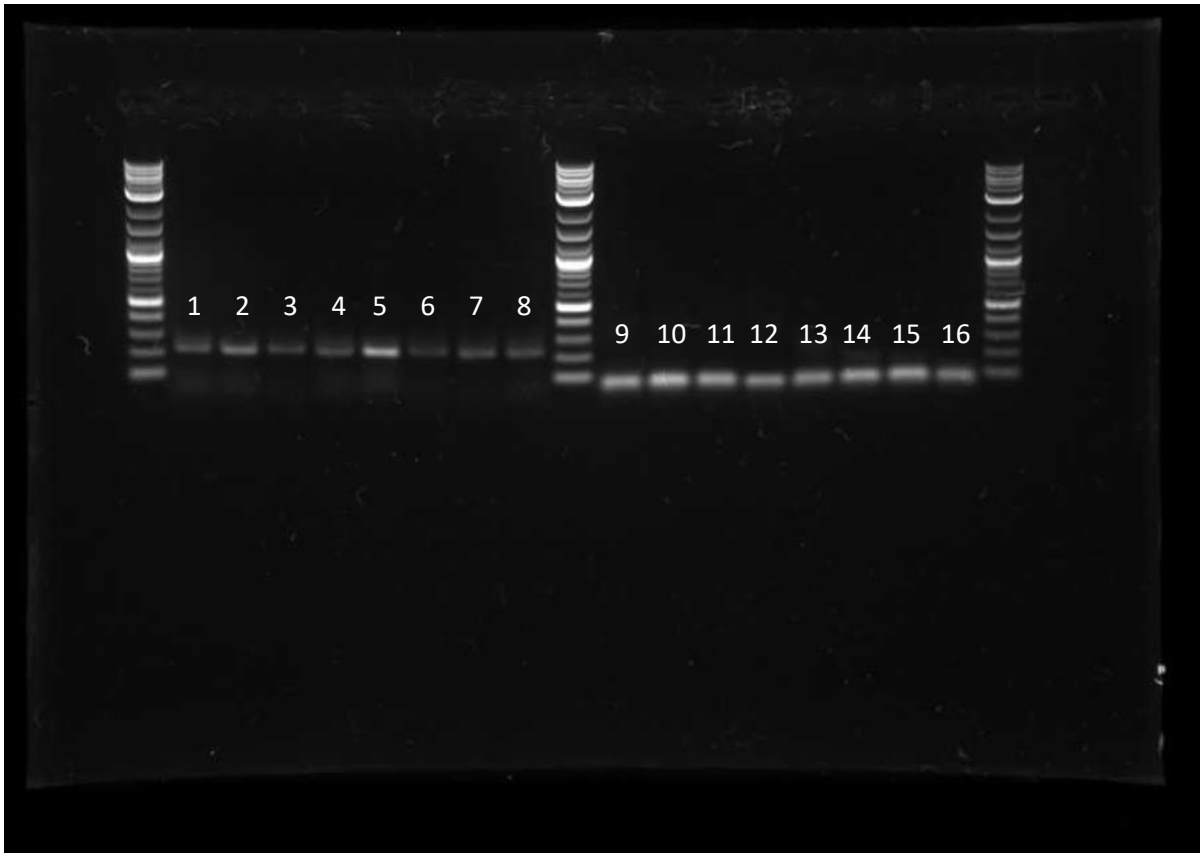
Another useful endeavour to advance the work completed in this project would be to use the method on bronchoalveolar lavage (BAL) samples and analyse this ‘gold-standard’ sample type in the same way that pleural fluid was used in this project. The availability of ethically approved BAL was a major caveat of this project, however a future study using BAL rather than pleural fluid could strengthen the results found in this study regarding its use on pulmonary clinical samples. If using BAL samples, it would be important to not use ‘blind-bronchoscopy’ to acquire the samples as this method is unable to provide definitive information on the location of the BAL within the lungs. It is recommended that samples are acquired through either electromagnetic navigational bronchoscopy (ENB) or fibreoptic bronchoscopy.

A further obstacle which could be worth future investigation is the relative quantity of host DNA compared to microbial DNA in each clinical sample. Human cells have significantly more genomic DNA than most bacterial species (largest sequenced bacterial genome = 14.7 million base pairs [91]; human genome = 14.8 billion base pairs [92]). This was managed during this project by only using lower concentrations (<1,000ng/µl) of each DNA template to reduce any impacts the host

DNA might have on the qPCR analysis. Human cells lyse far more readily than bacterial cells, so could be selectively lysed using low levels of detergent, however it must be noted that this would also remove all extracellular DNA from the sample. The issue of host DNA presence could also be mitigated by post-extraction procedures, such as the use of dPCR for analysis, or by performing statistical techniques such as decontam [93] to identify and remove contaminating sequences from metagenomics data after samples have been processed.

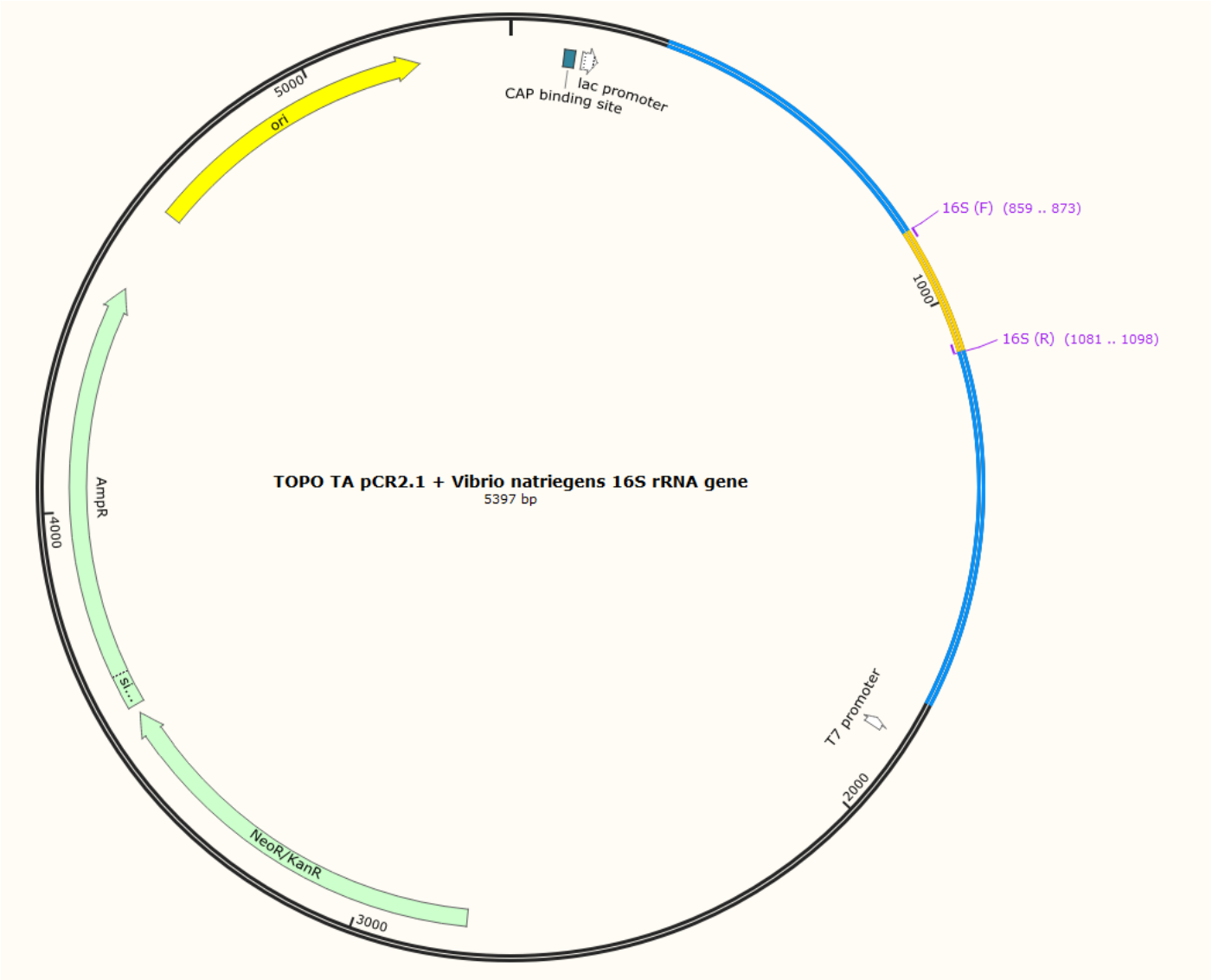
9. Appendix

9.1. Optimisation of primers for traditional PCR:



Temperature gradient PCR runs were carried out on the 16S rRNA primer set and the nuc_Hoegh primer set using JE2 genomic DNA as a template. This was done in order to determine the optimal annealing temperature (T_a) for each set of PCR primers. The bands on the left (1-8) show the PCR product from the run using the 16S rRNA primer set, the temperature gradient ranged from 50°C to 35°C. Band 5 was selected as the optimal T_a for the 16S rRNA primers (39.5°C). The bands on the right (9-16) show the PCR product from the run using the nuc_Hoegh primer set, the temperature gradient ranged from 60°C to 45°C. Band 10 was selected as the optimal T_a for the nuc_Hoegh primers (58.2°C). The gel was 1% agarose and electrophoresis conditions were: 300mA, 90V, 45 minutes. 5ul of PCR product was loaded per well and a 1kb ladder was used.

9.2. Plasmid map of 16S rRNA gene in TOPO TA pCR2.1



Plasmid map of the 16S rRNA gene from Vibrio natriegens in a TOPO TA pCR2.1 plasmid. The 16S rRNA gene is shown in blue with V4 region highlighted in yellow. The TOPO TA pCR2.1 vector is shown in black. Produced using SnapGene viewer [94].

9.3. Plasmid map of *S. aureus nuc* gene in pCDF-Duet-1

Plasmid map of the nuc gene from Staphylococcus aureus in a pCDF-Duet-1 plasmid. The nuc gene is shown in blue, the pCDF-Duet-1 vector is shown in black. Produced using SnapGene viewer [94].

9.4. *nuc* amplicon for production of p(*nuc*) standard

nuc amplicon: length = 749bp including the primers, restriction sites and ATAT repeats.

Sequence:

ATAT_GAATTC_CTAAAAAGAAAGAGGTGTTAGTTATGACAGAATACTTATTAAGTGCTGGCAT
ATGTATGGCAATCGTTTCAATATTACTTATAGGGATGGCTATCAGTAATGTTTCGAAAGGGCAA
TACGCAAAGAGGTTTTTCTATTTCGCTACTAGTTGTTTAGTGTTAACTTTAGTTGTAGTTTCAAG
TCTAAGTAGCTCAGCAAATGCATCACAAACAGATAATGGCGTAAATAGAAGTGGTTCTGAAGA
TCCAACAGTATATAGTGCAACTTCAACTAAAAAATTACATAAAGAACCTGCGACATTAATTAAA
GCGATTGATGGTGATACGGTTAAATTAATGTACAAAGGTCAACCAATGACATTCAGACTATTAT
TGGTTGATACACCTGAAACAAAGCATCCTAAAAAAGGTGTAGAGAAATATGGTCCTGAAGCAA
GTGCATTTACGAAAAAAATGGTAGAAAATGCAAAGAAAATTGAAGTCGAGTTTGACAAAGGTC
AAAGAACTGATAAATATGGACGTGGCTTAGCGTATATTTATGCTGATGGAAAAATGGTAAACG
AAGCTTTAGTTCGTCAAGGCTTGGCTAAAGTTGCTTATGTTTATAAACCTAACAATACACATGA
ACAACTTTTAAGAAAAAGTGAAGCACAAGCAAAAAAAGAGAAATTAAATATTTGGAGCGAAGAC
AACGCTGATTCAGGTCAATAATGCTCATTGTAAAAGTGTC_GGTACC_ATAT

The ATAT repeats are to assist the action of restriction enzymes. Restriction sequences are at the 5' end of each primer. A small region of DNA beyond the nuc gene was included at both the 5' and 3' ends to ensure the entirety of the nuc gene was included in the amplicon. The sequence for the nuc gene is shown in red.

9.5. Roche Kit: Lysostaphin vs Metapolyzyme

		DNA yield (ng/µl)				
		A	B	C	Avg. i.	Avg. ii.
Lysostaphin	Lys 1	69.0	68.5	68.5	68.67	71.67
	Lys 2	84.4	86.3	86.1	85.60	
	Lys 3	61.1	61.7	59.4	60.73	
MetaPolyzyme	Met 1	81.3	80.0	82.3	81.20	83.63
	Met 2	78.9	78.1	78.4	78.47	
	Met 3	92.1	91.9	89.7	91.23	

Comparison of DNA yields extracted from 500µl samples of JE2 liquid culture using the Roche High Pure PCR Template Preparation DNA extraction kit and either Lysostaphin or MetaPolyzyme for enzymatic cellular lysis. DNA yields were quantified using a NanoDrop spectrophotometer. The average readings for each biological repeat (Avg. i.) were used to produce an average for each enzymatic method (Avg. ii.). A two-tailed, unpaired t-test was performed which found that these two enzymatic methods do not produce significantly different yields (P=0.2227; T = 1.4423; N = 6; df = 4).

9.6. Fresh vs frozen

		DNA yield (ng/µl)				
		A	B	C	Avg. i.	Avg. ii.
Met + kit: Fresh	1	81.8	82.4	82.0	82.07	99.38
	2	91.2	91.7	92.5	91.80	
	3	123.9	119.1	129.8	124.27	
Met + kit: Frozen	1	94.7	99.2	96.9	96.93	91.78
	2	81.0	83.1	76.8	80.30	
	3	93.2	97.9	103.2	98.10	
BB +P/C: Fresh	1	175.6	199.8	197.4	190.93	192.83
	2	180.5	177.1	203.9	187.17	
	3	199.6	199.5	202.1	200.40	
BB + P/C: Frozen	1	190.3	185.4	184.3	186.67	191.03
	2	192.2	188.1	185.6	188.63	
	3	195.6	196.9	200.9	197.80	
Met + P/C: Fresh	1	186.5	190.2	188.7	188.47	203.87
	2	229.4	231.6	231.7	230.90	
	3	194.7	190.6	191.4	192.23	
Met + P/C: Frozen	1	127.0	114.6	115.2	118.93	123.14
	2	136.0	126.1	129.2	130.43	
	3	122.3	118.5	119.4	120.07	

Comparison of DNA yields extracted from either fresh or frozen 500µl samples of JE2 liquid culture using either MetaPolyzyme + the Roche High Pure PCR Template Preparation DNA extraction kit (Met+Kit), the Bead-beating + phenol/chloroform DNA extraction method (BB+P/C), or the MetaPolyzyme + phenol/chloroform DNA extraction method (Met+P/C). DNA yields were quantified using a NanoDrop spectrophotometer. Three readings were taken for each biological replicate and an average was taken (Avg. i.). An average was then taken from the three biological replicates per sample (Avg. ii.).

9.7. Comparison of kit vs P/C methods

		DNA yield (ng/µl)				
		A	B	C	Avg. i.	Avg. ii.
Met + Kit	1	81.3	80.0	82.3	81.20	83.63
	2	78.9	78.1	78.4	78.47	
	3	92.1	91.9	89.7	91.23	
BB + P/C	1	277.2	278.9	281.4	279.17	262.73
	2	2204.6	2152.7	2173.3	2176.87	
	3	250.2	251.9	236.8	246.30	
Met + P/C	1	130.4	132.8	137.0	133.40	130.01
	2	143.9	146.2	148.1	146.07	
	3	116.1	107.6	108.0	110.57	

Comparison of DNA yields extracted from 500µl samples of JE2 liquid culture using either MetaPolyzyme + the Roche High Pure PCR Template Preparation DNA extraction kit (Met+Kit), the Bead-beating + phenol/chloroform DNA extraction method (BB+P/C), or the MetaPolyzyme + phenol/chloroform DNA extraction method (Met+P/C). DNA yields were quantified using a NanoDrop spectrophotometer. An average was taken for each biological replicate (Avg. i.), these averages were then used to produce an average for each DNA extraction method (Avg. ii.). Results from the biological repeat BB+P/C 2 (red text in table) were omitted due to contamination of the sample.

It should also be noted that these experiments were carried out in parallel with the 'Roche Kit: Lysostaphin vs MetaPolyzyme' experiments. The three biological repeats which comprised the 'Met+Kit' samples in the 'Comparison of kit vs P/C methods' experiments were also the 'MetaPolyzyme' samples in the 'Roche Kit: Lysostaphin vs MetaPolyzyme' experiments shown in Appendix 9.5.

9.8. Comparison of methods using serially diluted culture

			DNA yield (ng/µl)				
			A	B	C	Avg. i.	Avg. ii.
Neat	met + kit	1	81.8	82.4	82.0	82.07	99.38
		2	91.2	91.7	92.5	91.80	
		3	123.9	119.1	129.8	124.27	
	BB + P/C	1	175.6	199.8	197.4	190.93	192.83
		2	180.5	177.1	203.9	187.17	
		3	199.6	199.5	202.1	200.40	
	met + P/C	1	186.5	190.2	188.7	188.47	203.87
		2	229.4	231.6	231.7	230.90	
		3	194.7	190.6	191.4	192.23	
1 in 10	met + kit	1	16.5	15.8	15.1	15.80	15.82
		2	16.5	14.7	19.2	16.80	
		3	14.8	15.3	14.5	14.87	
	BB + P/C	1	31.9	36.8	38.8	35.83	41.32
		2	43.1	41.6	44.2	42.97	
		3	45.4	45.2	44.9	45.17	
	met + P/C	1	17.1	17.3	19.9	18.10	17.17
		2	20.4	21.0	21.7	21.03	
		3	14.1	12.1	10.9	12.37	
1 in 100	met + kit	1	13.7	15.9	15.3	14.97	8.48
		2	2.7	0.3	0.2	1.07	
		3	4.8	11.1	12.3	9.40	
	BB + P/C	1	17.7	10.9	9.4	12.67	11.67
		2	8.8	7.7	7.4	7.97	
		3	14.2	15.2	13.7	14.37	
	met + P/C	1	13.0	15.4	19.5	15.97	10.53
		2	9.7	8.0	7.6	8.43	
		3	7.6	7.0	7.0	7.20	

Comparison of DNA yields extracted from 500µl samples of JE2 as either neat liquid culture, diluted 1 in 10, or diluted 1 in 100. Extractions were performed using either MetaPolyzyme + the Roche High Pure PCR Template Preparation DNA extraction kit (Met+Kit), the Bead-beating + phenol/chloroform DNA extraction method (BB+P/C), or the MetaPolyzyme + phenol/chloroform DNA extraction method (Met+P/C). DNA yields were quantified using a NanoDrop spectrophotometer. Three readings were taken for each biological replicate and an average taken (Avg. i.). An average was then taken from the three biological replicates per sample (Avg. ii.).

It should also be noted that these experiments were carried out in parallel with the 'Fresh vs frozen' experiments and that the 'neat' samples used in this experiment for all three DNA extraction methods are the same samples which comprised the 'fresh' samples for all three DNA extraction methods in the fresh vs frozen experiments shown in Appendix 9.6.

9.9. qPCR comparison of methods using serially diluted culture

		Cq values			
		1	2	3	Avg.
Met + kit	Neat	15.49	16.08	15.43	15.67
	1 in 10	23.88	22.68	23.18	23.25
	1 in 100	28.18	27.48	30.90	28.86
BB + P/C	Neat	14.52	13.55	12.60	13.56
	1 in 10	17.98	19.73	17.64	18.45
	1 in 100	26.58	25.59	26.23	26.13
Met + P/C	Neat	19.24	19.33	19.68	19.42
	1 in 10	24.67	25.13	26.97	25.59
	1 in 100	30.27	30.61	29.13	30.00
p(16S)	Neat	13.30	-	-	13.30
	1 in 10	17.40	-	-	17.40
	1 in 100	21.68	-	-	21.68
Water (NTC)	-	30.09	30.64	-	30.36

Cq values for the comparison of DNA extraction methods on serially diluted culture (neat, 1 in 10, and 1 in 100) using 16S rRNA primers. Samples included are three technical replicates for all dilutions of each of the three DNA extraction methods (Met+kit, BB+P/C, and Met+P/C). Also included is the p(16S) standard at the three dilutions (neat, 1 in 10, and 1 in 100), and two qPCR grade water non-template control (NTC) samples.

9.10. Standardisation of BB+P/C method: Phase lock gel

Sample			DNA Quantity (ng/µl)					Standard Deviation
			A	B	C	Biological repeat avg.	Sample type avg.	
JE2	PLG -	1	1846.0	1975.9	1949.0	1923.6	1425.01	396.15
		2	1248.7	1211.8	1419.7	1293.4		
		3	982.5	1021.7	1169.8	1058.0		
	PLG +	4	2053.9	2081.4	2259.0	2131.4	2014.81	115.81
		5	1979.0	1893.4	1903.4	1925.3		
		6	2012.9	1904.0	2046.3	1987.7		
Blank (TSB)	PLG -	7	19.4	17.0	25.8	20.7	19.54	4.12
		8	20.5	24.4	21.7	22.2		
		9	16.2	12.6	18.3	15.7		
	PLG +	10	28.2	17.3	20.0	21.8	18.82	4.57
		11	16.2	15.6	22.6	18.1		
		12	12.3	19.5	17.7	16.5		

DNA yields from phase log gel (PLG) tests to determine the effect of PLG on the variability of samples. Three technical repeats were performed for each biological sample (1-12), and three biological repeats were performed per sample type. Standard deviation was calculated to show variation of samples. DNA quantities were measured using a NanoDrop Spectrophotometer.

9.11. LoD experiments

9.11.1. LoD of qPCR stage only

Dilution of Plasmid Standard	Copy Number used	Cq Value					
		p(*nuc*)			p(16S)		
		A	B	Avg.	A	B	Avg
N	**2×10^7**	14.25	14.25	**14.25**	13.72	13.74	**13.73**
1×10^{-1}	**2×10^6**	17.23	17.56	**17.40**	17.20	17.88	**17.54**
1×10^{-2}	**2×10^5**	21.07	21.29	**21.18**	21.34	20.75	**21.05**
1×10^{-3}	**2×10^4**	23.96	24.03	**23.99**	23.79	24.15	**23.97**
1×10^{-4}	**2×10^3**	27.16	26.57	**26.87**	27.18	28.04	**27.61**
1×10^{-5}	**200**	29.12	29.02	**29.07**	29.42	30.23	**29.83**
1×10^{-6}	**20**	31.36	31.02	**31.19**	30.88	31.24	**31.06**
1×10^{-7}	**2**	-1	-1	**0.00**	31.60	32.11	**31.86**
Water NTC	**0**	-1	-1	**0.00**	32.15	31.70	**31.93**

Cq values from extended serial dilution using plasmid standards to determine the limit of detection (LoD) of the qPCR analysis stage with each primer set (nuc_qPCR and 16S rRNA). For each qPCR reaction, 2µl of each sample at $1x10^{(n)}$ was used as a DNA template.

9.11.2. CFU/ml of JE2 liquid culture for LoD experiments

Sample		CFU/ml				
Number	**Dilution**	**1**	**2**	**3**	**Avg.**	**Estimated CFU/ml**
1	**Neat**	TNTC	TNTC	TNTC	-	**~1,000,000,000 (1 x 10^9)**
2	**1 x 10^{-1}**	TNTC	TNTC	TNTC	-	**~100,000,000 (1 x 10^8)**
3	**1 x 10^{-2}**	TNTC	TNTC	TNTC	-	**~10,000,000 (1 x 10^7)**
4	**1 x 10^{-3}**	TNTC	TNTC	TNTC	-	**~1,000,000 (1 x 10^6)**
5	**1 x 10^{-4}**	TNTC	TNTC	TNTC	-	**~100,000 (1 x 10^5)**
6	**1 x 10^{-5}**	TNTC	TNTC	TNTC	-	**~10,000 (1 x 10^4)**
7	**1 x 10^{-6}**	1024	1052	1020	1032.00	**~1,000 (1 x 10^3)**
8	**1 x 10^{-7}**	110	117	130	119.00	**~100 (1 x 10^2)**
9	**1 x 10^{-8}**	9	10	11	10.00	**~10**
10	**1 x 10^{-9}**	0	0	1	0.333	**~1**
11	**1 x 10^{-10}**	0	0	0	0.00	**0**
12	**1 x 10^{-11}**	0	0	0	0.00	**0**
13	**1 x 10^{-12}**	0	0	0	0.00	**0**
14	**1 x 10^{-13}**	0	0	0	0.00	**0**
15	**1 x 10^{-14}**	0	0	0	0.00	**0**
B	**Blank (TSB)**	0	0	0	0.00	**0**

CFU/ml plate counts from a serial dilution of JE2 liquid culture used for LoD experiments. The average CFU/ml counts for later dilutions (1x10^{-6} to 1x10^{-8}) were extrapolated to predict the estimated CFU/ml of earlier dilutions in the series (N to 1x10^{-5}).

9.11.3. LoD of BB+P/C DNA extraction and qPCR stage combined

		Cq value					
		nuc			16S		
		A	B	Avg.	A	B	Avg.
CFU/ml of JE2	100,000,000	11.29	11.30	11.30	10.38	10.26	10.32
	10,000,000	16.05	15.85	15.95	15.12	15.29	15.21
	1,000,000	21.36	21.22	21.29	24.08	24.09	24.09
	100,000	25.12	25.25	25.19	27.95	28.08	28.02
	10,000	25.47	25.58	25.53	29.11	28.88	29.00
	1,000	27.67	27.71	27.69	30.57	31.12	30.85
	100	27.88	28.15	28.02	31.07	31.07	31.07
	10	27.19	27.12	27.16	29.71	29.30	29.51
	1	27.29	27.09	27.19	29.82	29.96	29.89
Blank (TSB)		29.18	29.01	29.10	33.14	35.07	34.11
NTC (Water)		-1	-1	0.00	30.33	30.18	30.26

Raw Cq values from qPCR to determine the limit of detection (LoD) of the BB+P/C and qPCR stages combined. qPCR was performed using both 16S rRNA primers and nuc_qPCR primers.

9.12. Clinical sample inhibition experiments: qPCR

9.12.1. PF and p(*nuc*):

	Cq value		
	A	B	Avg.
p(*nuc*)	15.29	15.42	15.35
p(*nuc*) + PF	15.47	16.09	15.78
PF	29.55	30.84	30.19
Water	30.37	30.72	30.54

Raw Cq values from qPCR test to determine impact of pleural fluid on the qPCR analysis stage using nuc_qPCR primers.

9.12.2. PF and 16S:

	Cq value		
	A	B	Avg.
p(16S)	18.72	17.13	17.92
p(16S) + PF	17.69	18.46	18.07
PF	24.33	24.61	24.47
Water	28.41	28.94	28.68

Raw Cq values from qPCR test to determine impact of pleural fluid on the qPCR analysis stage using 16S rRNA primers.

9.13. Clinical sample inhibition experiments: BB+P/C DNA extraction and qPCR analysis stages combined

9.13.1. JE2 spike – *nuc* primers

	Cq value						
	A	**B**	**C**	**D**	**E**	**F**	**Avg.**
PBS	26.00	26.23	25.65	25.48	25.97	25.67	25.84
PBS + JE2	15.06	15.35	15.66	15.72	16.43	16.53	15.79
PF	24.48	23.36	24.90	23.90	22.74	24.14	23.92
PF + JE2	15.42	15.53	18.76	18.93	13.27	13.52	15.90
p(*nuc*) 2x10^{7}	12.50	11.50	-	-	-	-	12.00
p(*nuc*) 2x10^{6}	13.75	14.36	-	-	-	-	14.05
p(*nuc*) 2x10^{5}	16.38	16.11	-	-	-	-	16.24
Water (NTC)	-1.00	-1.00	-1.00	-1.00	-	-	0.00

Raw Cq values from qPCR test to determine impact of pleural fluid on the BB+P/C extraction method when using nuc primers. A set of pleural fluid based samples and a set of PBS based samples were produced. Half of each set was spiked with S. aureus strain JE2. All samples were put through the BB+P/C DNA extraction protocol and analysed by qPCR using nuc_qPCR primers alongside the p(nuc) standard at three dilutions and a non-template control of qPCR grade water.

9.13.2. JE2 spike – 16S rRNA primers

	Cq value						
	A	**B**	**C**	**D**	**E**	**F**	**Avg.**
PBS	34.32	31.96	32.02	34.19	33.59	32.59	33.11
PBS + JE2	20.99	21.12	21.10	20.80	21.07	20.81	20.98
PF	25.72	24.57	26.23	24.33	24.52	26.10	25.25
PF + JE2	21.28	21.39	21.00	21.42	21.49	20.74	21.22
p(16S) 2x10^{7}	17.25	16.91	-	-	-	-	17.08
p(16S) 2x10^{6}	20.63	20.21	-	-	-	-	20.42
p(16S) 2x10^{5}	23.90	23.41	-	-	-	-	23.66
Water (NTC)	32.05	33.11	32.27	31.77	-	-	32.30

Raw Cq values from qPCR test to determine impact of pleural fluid on the BB+P/C extraction method when using 16S rRNA primers. A set of pleural fluid based samples and a set of PBS based samples were produced. Half of each set was spiked with S. aureus strain JE2. All samples were put through the BB+P/C DNA extraction protocol and analysed by qPCR using 16S rRNA primers alongside the p(16S) standard at three dilutions and a non-template control of qPCR grade water.

10. References

[1] National Health Service - NHS (2019). Pneumonia. [online] Available at: https://www.nhs.uk/conditions/pneumonia/ [Accessed 14 Mar. 2019].

[2] Khan, R., Petersen, F. and Shekhar, S., 2019. Commensal Bacteria: An Emerging Player in Defense Against Respiratory Pathogens. Frontiers in Immunology, 10.

[3] Wu, B. and Segal, L., 2018. The Lung Microbiome and Its Role in Pneumonia. Clinics in Chest Medicine, 39(4), pp.677-689.

[4] de Steenhuijsen Piters, W., Huijskens, E., Wyllie, A., Biesbroek, G., van den Bergh, M., Veenhoven, R., Wang, X., Trzciński, K., Bonten, M., Rossen, J., Sanders, E. and Bogaert, D., 2015. Dysbiosis of upper respiratory tract microbiota in elderly pneumonia patients. The ISME Journal, 10(1), pp.97-108.

[5] Dickson, R. and Huffnagle, G. (2015). The Lung Microbiome: New Principles for Respiratory Bacteriology in Health and Disease. PLOS Pathogens, 11(7), p.e1004923.

[6] Li, L., Mendis, N., Trigui, H., Oliver, J. and Faucher, S. (2014). The importance of the viable but non-culturable state in human bacterial pathogens. Frontiers in Microbiology, 5.

[7] Mangiaterra, G., Amiri, M., Di Cesare, A., Pasquaroli, S., Manso, E., Cirilli, N., Citterio, B., Vignaroli, C. and Biavasco, F. (2018). Detection of viable but non-culturable Pseudomonas aeruginosa in cystic fibrosis by qPCR: a validation study. BMC Infectious Diseases, 18(1).

[8] de Steenhuijsen Piters, W., Sanders, E. and Bogaert, D. (2015). The role of the local microbial ecosystem in respiratory health and disease. Philosophical Transactions of the Royal Society B: Biological Sciences, 370(1675), p.20140294.

[9] Hilty, M., Burke, C., Pedro, H., Cardenas, P., Bush, A., Bossley, C., Davies, J., Ervine, A., Poulter, L., Pachter, L., Moffatt, M. and Cookson, W., 2010. Disordered Microbial Communities in Asthmatic Airways. PLoS ONE, 5(1), p.e8578.

[10] Mathieu, E., Escribano-Vazquez, U., Descamps, D., Cherbuy, C., Langella, P., Riffault, S., Remot, A. and Thomas, M., 2018. Paradigms of Lung Microbiota Functions in Health and Disease, Particularly, in Asthma. Frontiers in Physiology, 9.

[11] O'Dwyer, D., Ashley, S., Wilke, C., Falkowski, N., Huffnagle, G., Moore, B. and Dickson, R. (2017). Lung Microbiota Are Altered in a Murine Model of Pulmonary Fibrosis and Are Necessary for Pathogenesis. American Journal of Respiratory and Critical Care Medicine, 195, p.A1003.

[12] Ashley, S., O'Dwyer, D., Falkowski, N., Moore, B., Huffnagle, G. and Dickson, R. (2017). Disruption of Lung Microbiota Precedes Peak Lung Injury in a Mouse Model of Acute Respiratory Distress Syndrome. American Journal of Respiratory and Critical Care Medicine, 195, p.A1005.

[13] Schott, C., Weigt, S., Turturice, B., Ascoli, C., Metwally, A., Andropolis, K., Perkins, D., Belperio, J. and Finn, P. (2017). The Microbiome as a Predictor for Bronchiolitis Obliterans in COPD Lung Transplant Patients. American Journal of Respiratory and Critical Care Medicine, 195, p.A1008.

[14] Dickson, R., Erb-Downward, J., Freeman, C., McCloskey, L., Beck, J., Huffnagle, G. and Curtis, J. (2015). Spatial Variation in the Healthy Human Lung Microbiome and the Adapted Island Model of Lung Biogeography. Annals of the American Thoracic Society, 12(6), pp.821-830.

[15] Levy, M., 2013. A New Definition of Ventilator-associated Pneumonia: Far from Perfect, Better Than Before. Annals of the American Thoracic Society, 10(6), pp.644-645.

[16] Zakharkina, T., Martin-Loeches, I., Matamoros, S., Povoa, P., Torres, A., Kastelijn, J., Hofstra, J., de Wever, B., de Jong, M., Schultz, M., Sterk, P., Artigas, A. and Bos, L. (2017). The dynamics of the pulmonary microbiome during mechanical ventilation in the intensive care unit and the association with occurrence of pneumonia. Thorax, 72(9), pp.803-810.

[17] Kim, E., Kim, E., Lee, S., Yang, S., Yoo, C., Kim, Y., Han, S. and Yim, J., 2012. Bacterial Yield from Quantitative Cultures of Bronchoalveolar Lavage Fluid in Patients with Pneumonia on Antimicrobial Therapy. The Korean Journal of Internal Medicine, 27(2), p.156.

[18] Erb-Downward, J., Thompson, D., Han, M., Freeman, C., McCloskey, L., Schmidt, L., Young, V., Toews, G., Curtis, J., Sundaram, B., Martinez, F. and Huffnagle, G. (2011).

Analysis of the Lung Microbiome in the “Healthy” Smoker and in COPD. PLoS ONE, 6(2), p.e16384.
[19] Hunter, J. (2006). Ventilator associated pneumonia. Postgraduate Medical Journal, 82(965), pp.172-178.
[20] Vincent, J. (1995). The prevalence of nosocomial infection in intensive care units in Europe. Results of the European Prevalence of Infection in Intensive Care (EPIC) Study. EPIC International Advisory Committee. JAMA: The Journal of the American Medical Association, 274(8), pp.639-644.
[21] Torres, A., Niederman, M., Chastre, J., Ewig, S., Fernandez-Vandellos, P., Hanberger, H., Kollef, M., Li Bassi, G., Luna, C., Martin-Loeches, I., Paiva, J., Read, R., Rigau, D., Timsit, J., Welte, T. and Wunderink, R. (2017). International ERS/ESICM/ESCMID/ALAT guidelines for the management of hospital-acquired pneumonia and ventilator-associated pneumonia. European Respiratory Journal, 50(3), p.1700582.
[22] National Health Service - NHS England (2016). NHS Chief Launches New Fast Track Funding So NHS Patients Get Treatment Innovations Faster. [online] Available at: https://www.england.nhs.uk/2016/06/treatment-innovations/.
[23] Koenig, S. and Truwit, J., 2006. Ventilator-Associated Pneumonia: Diagnosis, Treatment, and Prevention. Clinical Microbiology Reviews, 19(4), pp.637-657.
[24] Chastre, J. and Fagon, J. (2002). Ventilator-associated Pneumonia. American Journal of Respiratory and Critical Care Medicine, 165(7), pp.867-903.
[25] Shah, P. (2001). Staphylococcus aureus in lower respiratory infections: Clinical relevance of antimicrobial resistance☆. Seminars in Respiratory Infections, 16(3), pp.196-202.
[26] Santajit, S. and Indrawattana, N. (2016). Mechanisms of Antimicrobial Resistance in ESKAPE Pathogens. BioMed Research International, 2016, pp.1-8.
[27] Varshney, A., Kuzmicheva, G., Lin, J., Sunley, K., Bowling, R., Kwan, T., Mays, H., Rambhadran, A., Zhang, Y., Martin, R., Cavalier, M., Simard, J. and Shivaswamy, S. (2018). A natural human monoclonal antibody targeting Staphylococcus Protein A protects against Staphylococcus aureus bacteremia. PLOS ONE, 13(1), p.e0190537.
[28] Nash, A., Dalziel, R. and Fitzgerald, J. (2015). Mims' Pathogenesis of infectious disease. 6th ed. Amsterdam: Elsevier/AP, Academic Press is an imprint of Elsevier, p.18.
[29] Rello, J., Torres, A., Ricart, M., Valles, J., Gonzalez, J., Artigas, A. and Rodriguez-Roisin, R. (1994). Ventilator-associated pneumonia by Staphylococcus aureus. Comparison of methicillin-resistant and methicillin-sensitive episodes. American Journal of Respiratory and Critical Care Medicine, 150(6), pp.1545-1549.
[30] Kollef, M. and Micek, S. (2005). Staphylococcus Aureus Pneumonia. Chest, 128(3), pp.1093-1097.
[31] Towner, K., Talbot, D., Curran, R., Webster, C. and Humphreys, H. (1998). Development and evaluation of a PCR-based immunoassay for the rapid detection of methicillin-resistant Staphylococcus aureus. Journal of Medical Microbiology, 47(7), pp.607-613.
[32] Al-Talib, H., Yean, C., Al-Khateeb, A. and Ravichandran, M. (2013). Comparative Evaluation of Three Different Methods of Genomic DNA Extraction for Staphylococcus aureus. World Applied Sciences Journal, 21(3), pp.424-427.
[33] Ali, N., Rampazzo, R., Costa, A. and Krieger, M. (2017). Current Nucleic Acid Extraction Methods and Their Implications to Point-of-Care Diagnostics. BioMed Research International, 2017, pp.1-13.
[34] Liu, Y., Schulze-Makuch, D., de Vera, J., Cockell, C., Leya, T., Baqué, M. and Walther-Antonio, M. (2018). The Development of an Effective Bacterial Single-Cell Lysis Method Suitable for Whole Genome Amplification in Microfluidic Platforms. Micromachines, 9(8), p.367.
[35] Brown, S., Santa Maria, J. and Walker, S. (2013). Wall Teichoic Acids of Gram-Positive Bacteria. Annual Review of Microbiology, 67(1), pp.313-336.
[36] O'Riordan, K. and Lee, J. (2004). Staphylococcus aureus Capsular Polysaccharides. Clinical Microbiology Reviews, 17(1), pp.218-234.
[37] Archer, N., Mazaitis, M., Costerton, J., Leid, J., Powers, M. and Shirtliff, M. (2011). Staphylococcus aureus biofilms. Virulence, 2(5), pp.445-459.
[38] The Open University (2019). Understanding antibiotic resistance. [online] OpenLearn. Available at:

https://www.open.edu/openlearn/ocw/mod/oucontent/view.php?id=75460&section=4.1 [Accessed 17 Mar. 2019].
[39] Gordon, R. and Lowy, F. (2008). Pathogenesis of Methicillin-ResistantStaphylococcus aureusInfection. Clinical Infectious Diseases, 46(S5), pp.S350-S359.
[40] Hoegh, S., Skov, M., Boye, K., Worning, P., Jensen, T. and Kemp, M. (2014). Variations in the Staphylococcus aureus-specific nuc gene can potentially lead to misidentification of meticillin-susceptible and -resistant S. aureus. Journal of Medical Microbiology, 63(Pt_7), pp.1020-1022.
[41] Sigma-Aldrich, n.d. Pcdfduet™-1 DNA - Novagen. [online], Merck Millipore. Available at: https://www.merckmillipore.com/GB/en/product/pCDFDuet-1-DNA-Novagen,EMD_BIO-71340 [Accessed 6 December 2019].
[42] IUPAC (2019). International Union of Pure and Applied Chemistry (IUPAC) Codes. [online] Bioinformatics.org. Available at: https://www.bioinformatics.org/sms/iupac.html [Accessed 18 Mar. 2019].
[43] Life Technologies (2015). TOPO TA Cloning Kit User Manual. Invitrogen, p.29.
[44] Staroscik, A. (2004). Calculator for determining the number of copies of a template. [online] URI Genomics & Sequencing Center. Available at: https://cels.uri.edu/gsc/cndna.html [Accessed 18 Mar. 2019].
[45] Klenov, A. (n.d.). DIY Phase Separating Gel: Clean and Cheap!. [online] Bitesize Bio. Available at: https://bitesizebio.com/18944/diy-phase-separating-gel-clean-and-cheap/ [Accessed 18 Mar. 2019].
[46] Pipette Jockey (n.d.). DIY Phase Separating Gel (Phase lock) Redux: making the dense version. [online] Available at: http://pipettejockey.com/2017/06/21/diy-phase-separating-gel-phase-lock-redux-making-the-dense-version/ [Accessed 18 Mar. 2019].
[47] Quantabio (n.d.). 5PRIME Phase Lock Gel. [online] Quantabio.com. Available at: https://www.quantabio.com/phase-lock-gel [Accessed 18 Mar. 2019].
[48] Yang, B., Wang, Y. and Qian, P. (2016). Sensitivity and correlation of hypervariable regions in 16S rRNA genes in phylogenetic analysis. BMC Bioinformatics, 17(1).
[49] Postgate, J. and Hunter, J. (1961). On the Survival of Frozen Bacteria. Journal of General Microbiology, 26(3), pp.367-378.
[50] Meskauskas, A., n.d. Compute Reverse Complement Of Nucleotide Sequences. [online] Reverse-complement.com. Available at: http://reverse-complement.com/ [Accessed 24 October 2019].
[51] Corpet, F., 1988. Multiple sequence alignment with hierarchical clustering. Nucleic Acids Research, 16(22), pp.10881-10890.
[52] Fuchs, S., Mehlan, H., Bernhardt, J., Hennig, A., Michalik, S., Surmann, K., Pané-Farré, J., Giese, A., Weiss, S., Backert, L., Herbig, A., Nieselt, K., Hecker, M., Völker, U. and Mäder, U., 2018. Aureo Wiki–The repository of the Staphylococcus aureus research and annotation community. International Journal of Medical Microbiology, 308(6), pp.558-568.
[53] Ye, J., Coulouris, G., Zaretskaya, I., Cutcutache, I., Rozen, S. and Madden, T., 2012. Primer-BLAST: A tool to design target-specific primers for polymerase chain reaction. BMC Bioinformatics, 13(1).
[54] Salter, S., Cox, M., Turek, E., Calus, S., Cookson, W., Moffatt, M., Turner, P., Parkhill, J., Loman, N. and Walker, A., 2014. Reagent and laboratory contamination can critically impact sequence-based microbiome analyses. BMC Biology, 12(1).
[55] Dilhari, A., Sampath, A., Gunasekara, C., Fernando, N., Weerasekara, D., Sissons, C., McBain, A. and Weerasekera, M., 2017. Evaluation of the impact of six different DNA extraction methods for the representation of the microbial community associated with human chronic wound infections using a gel-based DNA profiling method. AMB Express, 7(1).
[56] Hearn, R. and Arblaster, K. (2010). DNA extraction techniques for use in education. Biochemistry and Molecular Biology Education, 38(3), pp.161-166.
[57] Browder, H., Zygmunt, W., Young, J. and Tavormina, P. (1965). Lysostaphin: Enzymatic mode of action. Biochemical and Biophysical Research Communications, 19(3), pp.383-389.

[58] Robinson, J., Hardman, J. and Sloan, G. (1979). Relationship between lysostaphin endopeptidase production and cell wall composition in Staphylococcus staphylolyticus. Journal of Bacteriology, 137(3), pp.1158-1164.
[59] Simões, M., Pereira, M. and Vieira, M. (2005). Action of a cationic surfactant on the activity and removal of bacterial biofilms formed under different flow regimes. Water Research, 39(2-3), pp.478-486.
[60] Jagielski, T., Gawor, J., Bakuła, Z., Zuchniewicz, K., Żak, I. and Gromadka, R. (2017). An optimized method for high quality DNA extraction from microalga Prototheca wickerhamii for genome sequencing. Plant Methods, 13(1).
[61] Inglis, P., Pappas, M., Resende, L. and Grattapaglia, D. (2018). Fast and inexpensive protocols for consistent extraction of high quality DNA and RNA from challenging plant and fungal samples for high-throughput SNP genotyping and sequencing applications. PLOS ONE, 13(10), p.e0206085.
[62] Bartram, A., Poon, C. and Neufeld, J. (2009). Nucleic acid contamination of glycogen used in nucleic acid precipitation and assessment of linear polyacrylamide as an alternative co-precipitant. BioTechniques, 47(6), pp.1019-1022.
[63] Ambion (2008). Linear Acrylamide. Applied Biosystems.
[64] Wang, G., Kang, S., Yang, H., Pai, S. and Chen, H., 2002. Removal of Dissolved Natural Organic Matter from Source Water with Alum Coagulation. Environmental Technology, 23(12), pp.1415-1423.
[65] MP Biomedicals, n.d. Lysing Matrix E, 2 Ml Tube. [online], mpbio. Available at: https://uk.mpbio.com/116914050-lysing-matrix-e-cf [Accessed 12 December 2019].
[66] Oswald, N., 2016. The Basics: How Phenol Extraction Of DNA Works. [online], Bitesize Bio. Available at: https://bitesizebio.com/384/the-basics-how-phenol-extraction-works/ [Accessed 12 December 2019].
[67] Plank, J. and New England Biolabs, 2010. Practical Application of Phenol/Chloroform Extraction. [online] Bitesize Bio. Available at: https://bitesizebio.com/3651/practical-application-of-phenolchloroform-extraction/ [Accessed 12 December 2019].
[68] Gaillard, C. and Strauss, F., 1990. Ethanol precipitation of DNA with linear polyacrylamide as carrier. Nucleic Acids Research, 18(2), pp.378-378.
[69] Liss, B., 2002. Improved quantitative real-time RT-PCR for expression profiling of individual cells. Nucleic Acids Research, 30(17), pp.89e-89.
[70] Sigma Aldrich, n.d. Genelute™-LPA Linear Polyacrylamide For Rapid Nucleic Acid Precipitation. [online], Merck. Available at: https://www.sigmaaldrich.com/Graphics/Supelco/objects/4800/4740.pdf) [Accessed 11 December 2019].
[71] Bloomfield, V., 1997. DNA condensation by multivalent cations. Biopolymers, 44(3), pp.269-282.
[72] Feng, B., Sosa, R., Mårtensson, A., Jiang, K., Tong, A., Dorfman, K., Takahashi, M., Lincoln, P., Bustamante, C., Westerlund, F. and Nordén, B., 2019. Hydrophobic catalysis and a potential biological role of DNA unstacking induced by environment effects. Proceedings of the National Academy of Sciences, 116(35), pp.17169-17174.
[73] O'Brien, L., Root, H., Wei, C., Jensen, D., Shabestary, N., De Meo, C. and Eder, D., 2015. M2+•EDTA Binding Affinities: A Modern Experiment in Thermodynamics for the Physical Chemistry Laboratory. Journal of Chemical Education, 92(9), pp.1547-1551.
[74] Higgins, C., 2009. Clinical Aspects Of Pleural Fluid Ph. [online], Acutecaretesting.org. Available at: https://acutecaretesting.org/en/articles/clinical-aspects-of-pleural-fluid-ph [Accessed 14 December 2019].
[75] Louca, S., Doebeli, M. and Parfrey, L., 2018. Correcting for 16S rRNA gene copy numbers in microbiome surveys remains an unsolved problem. Microbiome, 6(1).
[76] Stoddard, S., Smith, B., Hein, R., Roller, B. and Schmidt, T., 2014. rrnDB: improved tools for interpreting rRNA gene abundance in bacteria and archaea and a new foundation for future development. Nucleic Acids Research, 43(D1), pp.D593-D598.
[77] Větrovský, T. and Baldrian, P., 2013. The Variability of the 16S rRNA Gene in Bacterial Genomes and Its Consequences for Bacterial Community Analyses. PLoS ONE, 8(2), p.e57923.

[78] Popowicz, N., Lansley, S., Cheah, H., Kay, I., Carson, C., Waterer, G., Paton, J., Brown, J. and Lee, Y., 2017. Human pleural fluid is a potent growth medium for Streptococcus pneumoniae. PLOS ONE, 12(11), p.e0188833.
[79] UniProt (n.d.). nuc - Thermonuclease precursor - Staphylococcus aureus - nuc gene & protein. [online] Uniprot.org. Available at: https://www.uniprot.org/uniprot/P00644 [Accessed 18 Mar. 2019].
[80] Brakstad, O., Aasbakk, K. and Maeland, J. (1992). Detection of Staphylococcus aureus by polymerase chain reaction amplification of the nuc gene. Journal of Clinical Microbiology, 30(7), pp.1654-1660.
[81] Maes, N., Magdalena, J., Rottiers, S., De Gheldre, Y. and Struelens, M. (2002). Evaluation of a Triplex PCR Assay To Discriminate Staphylococcus aureus from Coagulase-Negative Staphylococci and Determine Methicillin Resistance from Blood Cultures. Journal of Clinical Microbiology, 40(4), pp.1514-1517.
[82] Louie, L., Goodfellow, J., Mathieu, P., Glatt, A., Louie, M. and Simor, A. (2002). Rapid Detection of Methicillin-Resistant Staphylococci from Blood Culture Bottles by Using a Multiplex PCR Assay. Journal of Clinical Microbiology, 40(8), pp.2786-2790.
[83] Tong, S., Schaumburg, F., Ellington, M., Corander, J., Pichon, B., Leendertz, F., Bentley, S., Parkhill, J., Holt, D., Peters, G. and Giffard, P., 2014. Novel staphylococcal species that form part of a Staphylococcus aureus-related complex: the non-pigmented Staphylococcus argenteus sp. nov. and the non-human primate-associated Staphylococcus schweitzeri sp. nov. International Journal of Systematic and Evolutionary Microbiology, 65(Pt 1), pp.15-22.
[84] Schmidt, T. and Relman, D. (1994). Phylogenetic identification of uncultured pathogens using ribosomal RNA sequences. Methods in Enzymology, pp.205-222.
[85] Illumina (2014). High-Speed, Multiplexed 16S Microbial Sequencing on the MiSeq® System.
[86] Klindworth, A., Pruesse, E., Schweer, T., Peplies, J., Quast, C., Horn, M. and Glöckner, F. (2012). Evaluation of general 16S ribosomal RNA gene PCR primers for classical and next-generation sequencing-based diversity studies. Nucleic Acids Research, 41(1).
[87] Choo, J., Leong, L. and Rogers, G. (2015). Sample storage conditions significantly influence faecal microbiome profiles. Scientific Reports, 5(1).
[88] RainDance (n.d.). RainDance ThunderBolts™ Next-Gen Sequencing Cancer Panel and RainDrop™ Digital PCR System Combine to Unlock Cancer Insights. [online] Raindancetech.com. Available at: http://raindancetech.com/raindance-thunderbolts-next-gen-sequencing-cancer-panel-and-raindrop-digital-pcr-system-combine-to-unlock-cancer-insights/ [Accessed 18 Mar. 2019].
[89] ZymoBIOMICS, n.d. Zymobiomics Microbial Community Standards. [online], ZYMO RESEARCH. Available at: https://www.zymoresearch.com/collections/zymobiomics-microbial-community-standards [Accessed 17 December 2019].
[90] Schloss, P., Westcott, S., Ryabin, T., Hall, J., Hartmann, M., Hollister, E., Lesniewski, R., Oakley, B., Parks, D., Robinson, C., Sahl, J., Stres, B., Thallinger, G., Van Horn, D. and Weber, C., 2009. Introducing mothur: Open-Source, Platform-Independent, Community-Supported Software for Describing and Comparing Microbial Communities. Applied and Environmental Microbiology, 75(23), pp.7537-7541.
[91] Land, M., Hauser, L., Jun, S., Nookaew, I., Leuze, M., Ahn, T., Karpinets, T., Lund, O., Kora, G., Wassenaar, T., Poudel, S. and Ussery, D. (2015). Insights from 20 years of bacterial genome sequencing. Functional & Integrative Genomics, 15(2), pp.141-161.
[92] Venter, J. et al*. (2001). The Sequence of the Human Genome. Science, 291(5507), pp.1304-1351. (*Human Genome Project Consortium).
[93] Davis, N., Proctor, D., Holmes, S., Relman, D. and Callahan, B., 2018. Simple statistical identification and removal of contaminant sequences in marker-gene and metagenomics data. Microbiome, 6(1).
[94] GSL Biotech. SnapGene software. Available at: https://www.snapgene.com/.